PRV

PATENT- OCH REGISTRERINGSVERKET
Patentavdelningen

PCT 00/01079

REC'D 02 AUG 2000
WIPO PCT

Intyg
Certificate

Härmed intygas att bifogade kopior överensstämmer med de handlingar som ursprungligen ingivits till Patent- och registreringsverket i nedannämnda ansökan.

This is to certify that the annexed is a true copy of the documents as originally filed with the Patent- and Registration Office in connection with the following patent application.

09/926596

(71) *Sökande* *Active Biotech AB, Lund SE*
Applicant (s)

(21) *Patentansökningsnummer* *9901961-4*
Patent application number

(86) *Ingivningsdatum* *1999-05-28*
Date of filing

Stockholm, 2000-07-24

För Patent- och registreringsverket
For the Patent- and Registration Office

A. Södervall
Anita Södervall

Avgift
Fee

PRIORITY DOCUMENT
SUBMITTED OR TRANSMITTED IN COMPLIANCE WITH RULE 17.1(a) OR (b)

BEST AVAILABLE COPY

PATENT- OCH REGISTRERINGSVERKET
Postadress/Adress Box 5055 S-102 42 STOCKHOLM
Telefon/Phone +46 8 782 25 00
Telex 17978 PATOREG S
Telefax +46 8 666 02 86

Field of the invention

The present invention is based on the finding that two fimbrial operons, the *saf* operon and the *tcf* operon, are specific for *Salmonella enterica* subspecies 1 bacteria and therefore have therapeutic use. Due to their specificity they can be used to provide vaccines against *Salmonella enterica* subspecies I as well as for detection of *Salmonella enterica* subspecies I. The *saf* operon is specific for all *Salmonella enterica* subspecies 1 bacteria and the *tcf* operon is specific for the serovar Typhi of *Salmonella enterica* subspecies 1.

All or part of the DNA-sequences of the genes encoding these proteins can be used as active agents in a vaccine against diseases caused by the *Salmonella enterica* subspecies I bacterial strains or for detection of said bacterial strains.

The present invention also relates to methods of isolating these fimbrial proteins, to antibodies directed against these proteins, and to a vaccine composition comprising these proteins or antibodies directed against these proteins for use in the treatment of infections caused by the *Salmonella spp.* The fimbrial proteins according to the invention or antibodies directed against them can be used for detection of *Salmonella spp.* bacteria.

Background of the invention

The members of genus *Salmonella spp* colonize and infect a wide range of different organisms. Many cause gastroenteritis and enteric fever in humans and domesticated animals while others are not associated with human disease (Saylers et al, 1994). The genus has been divided into two species, *Salmonella bongori* and *Salmonella enterica* where *enterica* can be further subdivided into seven subspecies, designated I, II, IIIa, IIIb, IV, VI, and VII (Reeves et al, 1989). *Salmonella enterica* subspecies I are preferentially associated with warm-blooded animals. Over 99% of all clinical *Salmonella* isolates are strains belonging to this subspecies, including serovars Typhimurium and Enteritidis, which are the major causes of *Salmonella* induced gastroenteritis in humans, and Typhi, the human specific causative organism of typhoid fever, the most severe form of human salmonellosis (Popoff et al, 1992).

Salmonella enterica subspecies I consists of over 1300 different serovars and is preferentially associated with warm-blooded animals (Baumler, 1997). Over 99% of all clinical *Salmonella* isolate are strains belonging to this subspecies,

including serovars Typhimurium and Enteritidis, which are the major causes of *Salmonella* induced gastroenteritis in humans, and Typhi, the human specific causative organism of typhoid fever, the most severe form of human salmonellosis (Popoff and Le Minor, 1992).

Today gastroenteritis and enteric fever can neither be prevented nor treated with good results. Typhoid fever is a substantial public health problem in developing countries. Each year 33 million people become ill and over 500 000 people die from this infection (American Institute of Medicine, 1986). Typhoid fever can be prevented by vaccination with attenuated bacteria, such as Ty21 and Vi vaccines and whole cell vaccines. Whole cell vaccines show a high incidence of side effects (Ashcroft et al, 1964, Yugoslav Typhoid commission, 1964). The vaccines consisting of attenuated strains of Salmonella typhi suffer from serious drawbacks. They must be administered as three or four spaced doses in order to stimulate protective immune responses (Levine et al, 1989). The treatment of Salmonella typhi with antibiotics is jeopardized since there are strains of Salmonella typhi that are resistant to chloramphenicol, ampicillin, and trimethoprim as well as ciprofloxacin (i.e. multidrug-resistant strains) (Rowe et al, 1997).

Accurate detection of *Salmonella enterica* subspecies I is today not possible. *Salmonella enterica* subspecies I can today only be detected by antibodies directed against surface proteins of *Salmonella enterica* subspecies I. The use of the sequences according to the invention makes it for the first time possible to rapidly and accurately determine the presence of *Salmonella enterica* subspecies I.

For many pathogenic bacteria, there is evidence that the filamentous surface protein structures called pili (fimbriae) are connected to the adhesion of the bacteria to the host cells. Pili proteins are very antigenic and are easily purified. Therefore pili preparations have been used as antigens for vaccination.

Summary of the invention

The invention relates to the objects as defined in the claims. The main object of the present invention is to provide two fimbrial proteins that are specific for *Salmonella enterica* subspecies I bacterial strains, the nucleotide sequences

encoding said proteins, as well as the corresponding amino acid sequences of for therapeutic and diagnostic use. Further are recombinant microorganisms provided, in which the nucleotide sequences according to the invention have been inserted.

An object of the present invention is to provide vaccine compositions for use in the treatment of *Salmonella enterica* infective strains, essentially pure Saf and Tcf fili protein of *Salmonella enterica* subspecies I and *Salmonella enterica* subspecies I serovar Typhi, respectively, as well as antibodies directed to these fili proteins.

A further object of the present invention is to provide the DNA sequences of the genes encoding the Saf and Tcf proteins. These sequences can be used for recombinant production of the proteins and for the preparations of vector vaccines against *Salmonella enterica* subspecies 1 and *Salmonella enterica* subspecies 1 serovar Typhi, respectively, as well as for diagnostic purposes.

Yet another object of the present invention to use purified Saf and Tcf protein from *Salmonella enterica* subspecies 1 bacteria for active or passive immunization of mammals, i.e. the proteins according to the invention can be comprised in a vaccine composition or be used to raise antibodies which can be comprised in a vaccine composition.

Finally, an object of the present invention is to provide a method for preventing or reducing the possibility of *Salmonella* infection of a mammal by administering the vaccines according to the invention. The invention may be more fully understood by reference to the following drawings and detailed description.

Brief description of the drawings

Figure 1.

Schematic representation of phage clones (named N10, D1, B1, F11) covering the entire cs7 insert of *Salmonella enterica* serovar Typhimurium strain SR χ 3181, i.e. comprising the *saf* fimbrial operon, i.e. *safA, B, C* and *D* (SEQ ID NO 1).

The clones were selected from partial *Eco* RI and *Bam*HI libraries in the Lambda Dash II vector. The cs7 insert is represented by a bold line. The extent

of respective phage insert is represented by horizontal bars. Name and size of the phage inserts are indicated on the left side of the figure.

Figure 2.

Schematic representation of the pTY52 cosmid comprising the *tcf*-operon (SEQ ID NO 2).

A *tcf* specific PCR fragment of 11105 bp was cloned into the Expand vector I cosmid (Roche). The insert is represented with a thick black line while vector sequences are represented with thin lines. Relevant restriction sites sequences are indicated. The position of the *tcf*-operon, i.e. *tcfA, B, C* and *D* (SEQ ID NO 2), is represented by a shaded arrow.

Figure 3.

The phylogenetic distribution of the identified genes on the cs7 insert was investigated using the well defined SARC collection, see Example 1.

Figure 4.

A 2 kb large internal *Eco*R I fragment was used as a probe in a Southern blot of the SARC collection, see Example 2.

Sequence listing

SEQ ID NO 1—DNA sequence of the genes encoding the precursor of the saf fimbrie unit of *Salmonella enterica* subspecies I.

SEQ ID NO 2—DNA sequence of the genes which encode the precursor of the tcf fimbrie unit of *Salmonella enterica* subspecies I serovar Typhi.

Deposit information

The phages carrying the inserted SEQ ID NO 1, i.e. phages clones B1, D1, F11 and N10 (see Figure 1) have been given the ECACC Accession numbers 99051922, 99051923, 99051924, and 99051925, respectively.

The cosmide carrying the inserted SEQ ID NO 2, i.e. cosmide pTY52 (see Figure 2) has been given the ECACC Accession number 99051926.

The depositions were made May 19, 1999.

Detailed description of the invention

The present invention is based on the finding that two fimbrial operons, the *saf* operon and the *tcf* operon, are specific for *Salmonella enterica* subspecies 1 bacteria. Due to their specificity they can be used to provide vaccines against *Salmonella enterica* subspecies I as well as detection methods for *Salmonella enterica* subspecies I. The *saf* operon is specific for all *Salmonella enterica*

subspecies 1 bacteria and the *tcf* operon is specific for the serovar typhi of *Salmonella enterica* subspecies 1, see Examples 1 & 2.

The main object of the invention relates to two fimbrial operons, the *saf* operon and the *tcf* operon, that are specific for *Salmonella enterica subspecies* 1 bacteria for terapeutic use.

Another object of the present invention is to provide vaccines against *Salmonella enterica* subspecies 1 induced gastroentritis, entric fever and typhoid fever.

A further object of the present invention is to provide methods to detect *Salmonella enterica* subspecies 1. The nucleotide sequences according to the invention are useful for constructing vectors for use as vaccines for insertion into attenuated bacteria in constructing a recombinant vaccine, for insertion into a viral vector in constructing a recombinant viral vaccine, or for direct inoculation as a nucleic acid vaccine. The pili proteins according to the invention, or antigenic fragments thereof, can be used for active immunization and antibodies directed against them can be used for passive immunization. All these applications of the sequences according to the invention are obtained by applying standard techniques known to the man ordinary skilled in the art.

Vaccines against *Salmonella enterica* subspecies I.
The genes encoding the saf and tcf fimbrial structures, or fragments thereof, may be incorporated into a bacterial or viral vaccine comprising recombinant bacteria, virus or fungi which are engineered to produce one or more immunogenic epitopes of the saf or tcf fimbrial structures. In addition, the genes encoding the saf and tcf fimbrial structures, or part thereof, operatively linked to regulatory elements, can be introduced directly as a nucleic acid vaccine, to elicit a protective immune response.

The proteins or antigenic fragment thereof, deduced from the nucleic acid sequences of the present invention are useful alone or in conventional vaccine mixtures in the vaccine compositions according to the invention. The proteins could be produced by chemical synthesis or recombinant expression according to conventional methods.

The proteins and peptides according to the invention can be obtained by using a host organism transformed or transfected with an expression vector obtained by insertion of a gene according to the invention, or part thereof, into a vector in a conventional manner. The vector which is used to construct the expression vector is not particularly limited, but specific examples include plasmids such as pET (Stratagen) and the like; and phages such as M13 (NEB), phage display libraries and the like. As expression regulatory sequence can among others T7 promotors and lac promotors be used.

An appropriate host to be transformed or transfected with the expression vector can be chosen among for example *E.-coli, Salmonella* or *Bacillus subtilus*. The transformed or transfected host is cultured and proliferated under suitable conditions.

After culturing, the peptides of the present invention may be purified by, for example, chromatography, precipitation, and/or density gradient centrifugation. The thus obtained peptides can be used as a vaccine or for the production of antibodies directed against said peptides, which can be used for passive immunization.

The purified preparation containing one or several proteins according to the invention, or parts thereof, is then formulated as a pharmaceutical composition, as for example a vaccine, or in a mixture with adjuvants. If desired the proteins are fragmented by standard chemical or enzymatic techniques to produce antigenic segments.

In formulating the vaccine compositions with the peptide or protein, alone or in various combinations, the immunogen is adjusted to an appropriate concentration and formulated with any suitable vaccine adjuvant. The immunogen may also be incorporated into liposomes, or conjugated to polysaccharides and/or other polymers for use in a vaccine formulation.

The different vaccines according to the present invention are administered to mammals in many different ways. These include intradermal, intramuscular, intraperitoneal, intravenous, subcutaneous, oral, and intranasal routes of administration. The vaccine doses will differ depending on circumstances such

as body weight, interferences with other administered medicaments etc. The upper limit is not critical unless the dose shows toxicity.

The peptides and proteins of the present invention are also useful to produce monoclonal or polyclonal antibodies for use in passive immunotherapy against *Salmonella enterica* subspecies 1. Human immunoglobulin is preferred. Antisera is obtained from individuals immunized with proteins or peptides according to the invention. The immunoglobulin fraction is then enriched, for example by immunoaffininty or affininty chromatography. Antibodies raised in a suitable mammal or in the patient to be treated, can subsequently be administered locally or topically, e.g. orally to the patient.

Detection of *Salmonella enterica* subspecies I in general.
The sequences according to the invention, or part thereof, or fragments hybridizing therewith, as well as the proteins according to the invention, or part thereof, and antibodies directed to said proteins, or antigenic fragments thereof, can be used in molecular diagnostic assays for the detection of *Salmonell enterica* subspecies I.

Nucleic acids having the nucleotide sequence according to the invention, or any nucleotide sequence hybridizing therewith can be used as a probe in nucleic acid hybridization assays for the detection of *Salmonella spp* in various tissues and body fluids of patients. The hybridization assay may be of any type including; Southern blots, Northern blots, colony blots.

PCR technology is the most preferred technology for detection according to the invention of *Salmonella enterica* subspecies 1. Primers of at least one selected from the 5' end and one from the 3' end can be used in PCR and other known tests to rapidly identify the presence of *Salmonella enterica* subspecies 1. This is according to conventional techniques.

The isolated and purified proteins and peptides of the invention can be used as diagnostics to measure an increase in serum titer of *Salmonella enterica* subspecies I-specific antibody since they bind strongly to these antibodies. A serum test sample can be screened for *Salmonella enterica* subspecies I by methods such as for example ELISA.

The invention further comprises the use of antibodies directed against the saf and tcf fimbrie structures for quantitative or qualitative determinations of the pili proteins of the invention, or fractions thereof, in cells, tissues or body fluids.

Detection of *Salmonella enterica* subspecies I by using nucleic acid hybridization technology

Nucleic acid hybridization technology can also be to detect *Salmonella enterica* subspecies 1 according to the invention. The nucleic acid probes chosen from parts of the sequences according to the invention can be either DNA or RNA. DNA sequences complementary to the sequences according to the invention can also be used. The binding of the probe to the target sequence, i.e. the hybridization, must not be perfect. Variations and mutations of the sequences according to the invention can be used as long as they hybridize good enough to detect *Salmonella enterica* subspecies I. The preferred length of the nucleic acid probes is about 10 to 400 nucleotides, most preferred not longer than 100 nucleotides.

The nucleotide probe is preferably chosen from the parts of the sequences that have the least variation. In the most preferred embodiments when screening for SEQ ID NO 1 (the *saf* operon, specific for *Salmonella enterica* subspecies 1) a nucleotide probe or PCR primer selected from nucleotides 37 368-37 868 should be avoided since this region is hypervariable.

The nucleic acid probes according to the invention are prepared by any conventional method such as organic synthesis, recombinant techniques, or isolation from genomic DNA.

The nucleic acid probes of the invention are labeled in a conventional manner to signal hybridization to target nucleic acid from *Salmonella enterica* subspecies I. The labeling may comprise a radiolabel, an enzyme, a bacterial label, a fluorescent label, an antibody, an antigen, a latex particle, an electron dense compound, or a light scattering particle.

The probes may be provided in a lyophilized form, to be reconstituted in a buffer appropriate for hybridization, or the probes may already be present in

such a buffer. The buffer may contain a suitable hybridization enhancer, detergent, carrier DNA, and a compound to increase the specificity.

Any conventional hybridization assay technique, such as dot blot hybridization, Southern blotting, sandwich hybridization, displacement hybridization and the like, can be used.

The target analyte polynucleotide of a microorganism may be in various media, most often in a biological, or physiological specimen. In most cases it is preferred to subject the specimen containing the target polynucleotide to any conventional extraction, purification, and/or isolation before conducting the analysis.

The sample containing the target analyte nucleotide sequence must often be treated to convert the DNA to a single-stranded form, which may be accomplished by a variety of conventional techniques, such as thermal or chemical techniques.

The following examples describe the isolation and specificity of the sequences according to the invention.

EXAMPLE 1

Identification and characterization of the *saf* operon.

The present inventors found, upon investigation of a 7 kb chromosomal region on centisome 7 originally isolated from the *S. typhimurium* strain SR-11 $_k$3181, a region that exhibits many of the traits that define a pathogenicity island. It has a lower G+C composition than the average composition of the *Salmonella* genome and includes many sequences related to different mobile genetic elements. The region is not present in *E.coli* K12, and the *Salmonella* specific DNA is inserted between the tRNA gene *aspV* and the stop codon of *yafV*, a hypothetical protein upstream of the *yafH* gene at 5 min in the *E.coli* chromosome. This *Salmonella* specific insert encodes proteins creating adhesive structures and other virulence factors. Sequencing revealed genes encoding a new fimbrial operon that they designated *Salmonella* Atypical Fimbriae (*saf*), due to its relatedness to a subgroup of adhesive structures forming thin atypical fimbriae or non-fimbrial adhesins.

The *saf* operon consists of four contiguous genes, *safA*, *safB*, *safC* and *safD* that encode fimbrial subunit, periplasmic chaperone, outer membrane usher protein and alternative fimbrial subunit, respectively. The genes *safA, B, C* and *safD* encode putative proteins of 166, 244, 836 and 156 amino acids, respectively. Analyzes of clinical *Salmonella* isolates showed that DNA of 195 out of 198 clinical isolates belonging to *S. enterica* subspecies I hybridized with *safB* and *safC*, i.e. these sequences are common to more than 99% of the known *Salmonella enterica* subspecies 1 bacteria. The inventors showed that 58% of these clinical isolates carry the *safA*, see Table 1.

Table 1. The prevalence of the *saf* genes in clinical *Salmonella* isolates.

Serovar	*safA*	*safB*	*safC*	# isolates
S. adelaide	-	+	+	1
S. agona	+	+	+	6
S. anatum	-	+	+	3
S. bareilly	+	+	+	3
S. blockley	+	+	+	3
S. bovismorbificans	-	+	+	5
S. braenderup	-	+	+	4
S. brandenburg	+	+	+	1
S. bredeney	+/-	+	+	15
S. chester	+	+	+	1
S. colindale	-	+	+	1
S. derby	-	+	+	1
S. dublin	-	+	+	1
S. eastbourne	+	+	+	2
S. emek	+	+	+	1
S. enteritidis	-	+	+	8
S. give	-	+	+	1
S. goettingen	+	+	+	1
S. haardt	-	+	+	1
S. hadar	+	+	+	16
S. heidelberg	-	+	+	1
S. hvittingfoss	+	+	+	5
S. infantis	-/+	+	+	6
S. java	-	+	+	1
S. javiana	-	+	+	1
S. kottbus	-	+	+	1
S. livingstone	-	+	+	1
S. london	+	+	+	1
S. maastricht	+	+	+	2
S. mbandaka	-	-	-	3
S. montevideo	+	+	+	1
S. muenster	-	+	+	1
S. newport	+	+	+	2
S. ohio	+	+	+	1
S. oranienburg	+	+	+	2
S. panama	+	+	+	3
S. potsdam	+	+	+	1
S. rissen	-	-	-	1
S. saarbrucken	-	+	+	1
S. saint paul	+	+	+	3
S. schwartzengrund	-	+	+	1
S. singapore	+	+	+	1
S. stanley	+	+	+	5
S. subsp I 4.5,12:-:-	+	+	+	2
S. subsp I 4.5,12:b:-	-	+	+	1
S. subsp I 4.5,12:i:-	+	+	+	1
S. subsp I spont	-	+	+	1
S. tennessee	+	+	+	2
S. thompson	-	+	+	1
S. typhi	-	+	+	1
S. typhimurium	+	+	+	27
S. virchow	+	+	+	7
S. weltervreden	-	+	+	1
S. worthington	-	-	-	2
S. subsp III	-	-	-	1

The phylogenetic distribution of the identified genes on the cs7 insert was investigated using the well defined SARC collection, which showed that the presence of the *safA, safB, safC* and *safD* genes is restricted to *S. enterica* subspecies I (Fig. 3). This region is hence the first subspecies I specific genetic region to be identified with a broad distribution within the subspecies. Since the serovars of subspecies I constitute over 99% of human salmonellosis and are preferentially associated with warm blooded animals, it implicates a role for the *saf* adhesive organelle in the colonization of these organisms.

EXAMPLE 2

Identification and characterization of the *tcf* operon.

The present inventors found that *Salmonella enterica* subspecies I serovar Typhi contains DNA encoding an additional fimbrial operon, the tcf operon, in the *sinR-pagN* intergenic region. Southern blot analysis revealed a markedly different restriction pattern in *S. enterica* serovar Typhi than the other subspecies I isolates, suggesting that the *saf-sin* region in serovar Typhi might carry additional DNA relative to serovar Typhimurium strains. A PCR reaction (using a kit from Roche) was therefore performed using a *sinR* (5'-GTA AAT CGC TTA GTC GCC-3') specific forward primer and a *pagN* (5'-TCA ACT CAA CCT TCA GCC-3') specific reverse primer.

This primer pair produced, as expected, a product of 2 kb in serovar Typhimurium from the SARC collection, while from serovar Typhi the product was 10 kb. Thus, the neighboring *sinR* and *pagN* genes in serovar Typhimurium strains are separated by approximately 8 kb in serovar Typhi.

The Typhi specific PCR product was purified, digested partially with *EcoRI* and sub-cloned into pUC18 forming a set of overlapping clones. Sequencing of the clones revealed a putative fimbrial operon designated *tcf* for Typhi Colonizing Factor. Four ORFs, *tcfA,B,C,D*, have been identified with putative proteins having significant homology to CooB (38% identical over 192 aa), CooA (37% identical over 170 aa), CooC (34% identical over 872 aa) and CooD (31% identical over 272 aa), respectively. The Coo proteins are involved in the biosynthesis of the CS1 colonizing factor antigens of enterotoxigenic *E.coli* (Fig. 4) (Froehlich et al., 1994). The peptide of the *tcfB* ORF is also homologous to the CblA major fimbrial subunit protein (45% identical over 154 aa) of the cable

type II pili of the cystic fibrosis-associated *Burkholderia cepacia*(Sajjan et al., 1995). Down-stream of the *tcf*-operon two ORFs were identified with the same transcriptional orientation as the *tcf* genes. The first was designated *tinR* for Typhi insert regulator because it is homologous (33% identical over 144 aa) to AzlB of *Bacillus subtilis*, a member of the Lrp/AsnC family of transcriptional regulators (Belitsky et al., 1997). *tinR* is followed by an ORF (*tioA* for Typhi insert orf) encoding a putative protein of 205 amino acids with no significant homologies to anything in the DDBJ/EMBL/GenBank databases. The above sequence from *Salmonella enterica* serovar Typhi strain RKS 3333 and the *tcf* region of the incomplete genome sequence from serovar Typhi strain CT18 (http:// www.sanger.ac.uk) are 99% identical over the total length of the investigated region in concordance with the clonal nature of the serovar .

A 2 kb large internal *Eco*R I fragment was used as a probe in a Southern blot of the SARC collection. This blot shows that *Salmonella enterica* subspecies I serovar Typhi (SARC2) is the only strain in the collection possessing DNA hybridizing to this fragment (Fig. 4).

SEQUENCE LISTING NO. 1

<110> Folkesson, Anders

<120> The complete sequence of the cs7 insert in Salmonella enteric serovar Typhimurium

<130> Complete sequence of the cs7 insert

<140>
<141>

<160> 5

<170> PatentIn Ver. 2.1

<210> 1
<211> 46870
<212> DNA
<213> Salmonella typhimurium

<220>
<221> CDS
<222> (37368)..(37868)
<223> safA putative fimbrial subunit

<220>
<221> CDS
<222> (37952)..(38689)
<223> safB putative periplasmic chaperone

<220>
<221> CDS
<222> (38713)..(41223)
<223> safC putative outer membrane usher

<220>
<221> CDS
<222> (41245)..(41715)
<223> safD putative fimbrial subunit

<400> 1
gatacaaatc tcagggtgtt tttatacatc ctgtgaagta aaaaaaaccg tatcactgta 60

aaagggatac ggtttttttt cgtcttcaag aagttccacc gtctatcgtg gaatctggcg 120

caaatgggcc tacgcctgga tgacgaacag gatattaccg ccacttcttt cactgtcatg 180

gctattttga tcccactgac atttaaggcg cggcctcatg gcggtgctta accgggatcg 240

ggacatgttc agcgcagaag cagactgcgt aatgttgata tcactcagat aattacggag 300

aaccgccaga catgcgcatc atcactccag ggcatcccac ttctccagca actccaccgg 360

gatctcattg atcacctccg agaaccgttt tcccaccagt ctttcagcct ggcgtaacag 420

tgggatggtc gggctactgg gttcactgct ctcaaaccag cgacgaatgc accgcaggcg 480

ttccagcgca tcattacgat cgcgaattgg ccccggttca gcattgtgtg ggagggatat 540

accgggtggc gcgttgacag gcggcacatc ctctgcccgt gccggaacgg cattatccat 600

aaccgtgcct gcggggtctg gcactggcgc tggcgacgga gggatttccg gtgtggttgt 660

ctgtaccacg tcaggcagca gcgctaacaa ctgccgcagg cgtgaaaaat ctggagccag 720

atcgcctaat gtctcacgcg cccacacttg caggcgttct gcgctttcct gcgcttcgcg 780

aaacgccgcc agcggtaatg ccccacgggc ttcgaggtcg gccagttgct ggcggacgga 840

ttcgggggcc agcgcatcag ccgggcgggg agcggataac gcgcgctcca catcccgtac 900

ctgtaggcgc agggcggcac tgtttgacag cgtaataccg cgaatgtccg ccatcacgcc 960

ttcgtgatcc agcagcgccg ccagggcatt actgcgcgcc agcgacgcat cttcaataga 1020

ttctgcggat gcgccttcgc cggtaagcag ctgtggatgg agtgcatcag accagattac 1080

gctcagttct gccagttgtg tgagcatttc cgccagtccc tgtgcgccag cctgctggat 1140

gcgactgcgc agcagcagga tcaacacccg gatatcctta ctgcgcgtga gcagacggcg 1200

cgcatcacgt tcaatttccg gccagttcac ggcttccggc gtactgacaa aatcaccata 1260

ttgcgcctcc gcctgggggg cggcgcggct gaacagcagc aggtattccg gatcgtactc 1320

cggatcgggg ccgcagggtt gttccgcgct taccggttta gtcagggaca tgtccatatc 1380

attactctca gtgggttaag ccgtgttcag gttcaaaaat catgccggtc accggacggt 1440

catgcatggc tttacctgcc caggccgtcc agccaagccg ctgcgcgccg ccgatcaccg 1500

caggcggtgc ggcatgaggc ttcaactgca gctcaatttc ccagcaatat tcaaaaccga 1560

tgaacgtacg caccagttct gtcagcacgg gcagattgtt gccgccgggc agaaaacgca 1620

ggtaatcctc cagcgtgagc gggccgataa ttagccggaa tttgtactgc atatccggta 1680

cggcctggcc gataagcgcc ccgttgccca gcaccgaaga ctgcgcgggc gtgcccagac 1740

gggtgatttc atcgttcgcc accgttatcc agtgcagggc gaattctttc accgcaaaag 1800

gtacgctgaa atagtgcgcc agcgtggcgg ccagcccgtc aggattgcgc gattcgcgta 1860

ccagatgggc ggaggctgcc aggcgaacat gatctgacag cgggctttcg gcgctttccc 1920

gtagatcctg cccgctgaga ctggcgatat aaaacgcaaa acggtcgtgt tccggtttgt 1980

ccagcccgcc accagcggac tgggcgctgc gccatgcctg ccagaactgc gtcagccagc 2040

ggtggtgaaa aatattggaa aaatgaacca gcgtcggatc gtgacgactc tctgagcggt 2100

tcagtgccag ctcggtatag tgcagcggca acgggccgtt tggcccccat agtccgaggc 2160

tgtacaggct caggtgcagg cgtccatcct gccagctgac ctgggcgatt tcccgtggcg 2220

caaaggtcat cgtcggcgtc tgtcccagac ggaatttttc catccgtggc ttgccagata 2280

cttgcctgcc ggagtatcac agagctgggc atccacgcgc cgcatcaggt tcaggaatcc 2340

gtaacgccag ggggttttct gcgcctggtt gagagcgccg gtcatacgct cccccgttgc 2400

ccggttctga ccggccaggt catgacatgc ccgcgttgca tcgagtgcag cgtcatctgc 2460

gagaaggtat taatggaaac atggcgggca atatagtgtt ccagcaccag accgaacagg 2520

taaggactga taccggaaaa tccttcttcg tccacggtca gttcgcaact gacgccccgg 2580

ccatagacca acaggccgga gccgggcagg cggcgggtca ccggggtggt tttgcagcca 2640

atcaggctgc gcacctggcg cgactgcggg ctgtcgtgag ccgggataaa cagattcagc 2700

agatcgcgca gcgcctggcc gccggtgcgg tgatccagat cggccagcgg cagataatta 2760

aacgacaact gccggatcag ccgccaggcc atttcgcgtt cagccagcgg cggctgcggc 2820

gggcgcggcg gtctgataag acccacgccc gccaccggaa tcgccgcatc tacggtcaga 2880

tcatcccggc cattacgtgg aataaggcag ggcagatcgc ggttagtcac cattgccgtg 2940

acggtgatat ggcgcagatt ttccgggtag ggcgcttcat gctgatcaac cagcgagagg 3000

aagacttccg agccggtata gggggttcgg gtgccatagc ggcgggcgtt ttctgacgag 3060

cggcgcggtt cacgacgcag tgaaaaataa cgcccgtggt tgccttcgtc attattacgg 3120

gtgtgataca gcggacgaaa aatcatcttc cgtgtggttt ccgcttccag gccttcgact 3180

tcctgaacag aaaacacctc gtaatccagc ggacgggtac gatccaccac cagatgctgt 3240

tccgtcacgc tgtgagtgac ttcaatccgg gtggtggtgc gaggaagcag gttgatcacc 3300

ggcgtacaga acaggctgaa ctgtgcagcg tccgtctgat gaatcagcca gtccggcggc 3360

aggcggttaa gcagtatgac aatttccgcc acattgccct gcacctttg taacccggca 3420

gacaatccgg tcggggtgaa gaagtaaaac cgttccggac aggcgaaaaa ttcatgcagc 3480

agattatggc cgtgaaacac gttccaggcg agcggtagca gcccctgccc tggctccagc 3540

ccttcgtgcg ccaccgggtg ttgaagattc acattcagtt cgccgtcaaa gtgaccgggt 3600

tcaccggcca gtgtggcgac ggcgctggta tgtagcagct caaacaggtg tgacgcaatg 3660

cgttcttcgc cgcagaggta aaagggcagc cgtgccggac cggccagctc gctgaaagtc 3720

agttccccga aggttcgcag ggtgatgcgc aatgccccgg cgacatgaat attaggcggc 3780

agatagcggt gcagggcggg catatccggc ggcgcggcgg tcaggcgtac ctcctcgatg 3840

gacagcggcc acagcgtgac gtcctggctg ctgcgaaact ggcaggcggt attttcgcct 3900

tccgggatgg gggaaacgaa cgcggtatcg cgcggcacgg tgacccctttt cgccaggtcg 3960

ccttcctgcg tatcgggata cagctttacc actgccatcg atggcgtggg ggtgacgtaa 4020

ttggggctga cgacttccag taaccgctgt gtgaagcggg gaaactcggc gtcaattttt 4080

agctgagtgc gggcgctcag aaagctgaac gcctcgatca tgcgttccac atacgggtcg 4140

gcaatatcgg ttccctgcat ccccagtcgg gcggcaattt tgggatggag ggtggcgaac 4200

tcagcaccgg tctcccgcag gtagctcagt tcgcggttgt aatactccag tagccgtgga 4260

tccatgaata atgccctgta ttaaaagaac gtaatgcggc tcagctccat atccagcgcg 4320

ctgcgtacca gaaactccgt gggatacggc tgcgtcagaa tttgtccgcg aatttcaaac 4380

tgtagcgtgt tatagctacc ctgccgggtt ttatcgagca agggcgtgac cctgagtgtg 4440

gcggcgttca gccggggttc gaaacggata atggcgcgcc ggatcgcctc gctgatatcg 4500

tcccacttat gctcatacat aaagctgccc gccagcggcg gcaggccata gttgagcact 4560

gacgccgccg cctgcggata gcgccgggcg tcgatgtcac cctcgtggct aatggtattg 4620

agcaaaaagg agagatcccg gcgaatgatc tccttcagtt gtaccggcgt gacgctgata 4680

tcccggtcaa ttttctgata cggagcattg tcacacagcc gatcaaacag cgtgggcagc 4740

aggtggttag cgggtgtaaa acgagacgtg ctcatgcgcc atcgttttcc tgagcatgaa 4800

aggtacaatg ggccatgtcc agcaggctga tatcgccgtg gctggtcagc cacactttct 4860

gccccagcgc ccgcacggtg gtttcgccgg ggccgtcctg ccaggcggtt tccctgcaca 4920

gacgcagggc gtcggatgca ctttccgaac cgctgtaacg ggtaaagagc caggcgccgt 4980

gcgtatcgcc attcaccagg gtgatattaa cgggtttcca cagcagatca gtcaggcgcg 5040

tcggttgcgg cgattccagc gagcgtattt gcgaaaacgg cagccagata tacacgccgc 5100

cggtgactag ctcaagtacc gggccaaggc gggaatcgct gtcgctcgcc cagtcaaatg 5160

cgccgccgtt ccactgcccg cccgtgtctg ttatggcttc cagtgcggta ttacggtgtt 5220

tatcaacctc accggtatcg tcatgacagg cgagtgccgc cagcagtgac tccacccaaa 5280

cgggctgcgg cagaagaaaa ccgggtcgtt gttcaccctg aaaaacggtg tggcggaaca 5340

tttcgcagcg aaccagctcc cggtacagcc gggcctcctg ggtataattg gcctccatcc 5400

tggcgcatag ctgaagctgg tgtagcgccc gcgaccagtc tccggccaca cacagcaact 5460

gaaacaggct gtggcggcag agcgctttcg ccggattttc ccgaacctgc tgctccgcca 5520

tctgaatccc ctccgcaata gagtattcct gtatcagcgc ggacagggta gcaggaagcg 5580

tgtcagtttt tttcatgggc ggtatttcca tttttctgtg tcggagtgat tcggtagtgg 5640

ctatcgatgc caataatacg atgctcgcgg cgggtcagat cgggcagtat ctcgctatgt 5700

gcggttttcc cccccagttc cggcgagagc agttgcagga tatcgggcat tgattccatc 5760

gccagccagt gcatttcacc ctcaccggtc gtgtccagcg tatccagaat ggcatcaata 5820

ccgggcgcgc ctgccaccat atcctgtaac gtgtcggtgt cgccctttt atcgtataac 5880

gacgtcaaat cctgggtgac agcgtcttca tttcggggaa acggtttagc acggaggggc 5940

ttttcctgcc cggagggccg caatgcctgc tgatattcct ggtaaagctg gtgtagtgtg 6000

ctatccgctt catggccagt gtaagaggat gcggtttccc ggacgggaat aatatcgaac 6060

gggttctgcg gattgagctg ttgttgttta aaccagtcca gatccaggta ctcagcaacg 6120

gattcagatg actgaaccaa ctgaggaaac gaaacatcgg gacgggactc atcattggtg 6180

cgggcaaggc accaggatga cagaccccat tcaatcgtat ctccatcatt cagccgcata 6240

cggtgatgag gttccatcac ctgctcattg acggcacagc acagttcatc agagtgatta 6300

acaatccacc acgcctcttc gtggcgaaca agactgagcg cgatatcgtc tttcccatgt 6360

tgctcaaaaa gcggaatgta gggggcggcg gcagtaaata cgatgtatgt ccctgcggga 6420

tactgtgcgt tatccgttat accctgcgac tgtattttcc gtaattccca catgtcttgt 6480

ttcattatgt cccttaaatg tatttatttt tctggaggaa acgtttaggg agttttaatt 6540

cataaataat tttaactaaa tttataggga gtcattattg atgacaccct ttttattatt 6600

ttcgtcgtga atgcattgtt gtatgcatag atgtcttttt tgaaatatta tttcttttaa 6660

ttctgcaatt gcgattttta tcttgctgtt cattgattat ttaatattaa cggcttagtt 6720

gctaattagt ccctgatctg tatcattgtt ttgtttcgat attttttcga ggctatcaat 6780

aaagaattgt ttatatattt atatgcattg atgcattatt tttatgaatt tttatgtcac 6840

aaggcataac acatggaaac tcctgtttca cgcagtgcgt tgtatggaaa actggccggc 6900

ccactattcc ggtcgctgga atcggcaacg gcattttgca aactacgctc taatccctgg 6960

gttgagctga ctcactggct gcaccagtta acacagcagc ccgataacga tattctccac 7020

gttcttcggc attaccagat ccctctttct gatgtggaga aagcgttact ccggcaactg 7080

gatatgctgc ccgccggggc cagcgccatt agtgattttt ctcaccatat cgatctcagc 7140

gttgaaaagg cctggatgct ggcgagcgtc cgttacggcg ataacaaaat tcgcagcggc 7200

tggttgctgc tggccttgtt gaccacgcca gaactgcgtc gggtactgag cagtatctgc 7260

gcgccgctgg ccacgcttcc ggttgatgaa ctgacggaaa tactgccctc gttgatcgaa 7320

acatcgccgg aagcgcagga gcgcccttac gacggctccg gcctggcatc agccattccc 7380

ggtgaaagca gtcaggcgat tcccaacggc gggcaggacg gtaaatccgc gctggcaaaa 7440

tactgtcagg acatgacggc acaggcgcgc gacggcaaaa tcgacccggt gacggggcgt 7500

gagcatgaaa tccgcaccat gacggatatt ctgctgcgcc gtcgccagaa taatccacta 7560

ctgactggtg aggcgggcgt cgggaaaacg gcggtcgtcg aaggttttgc cctcgcgatt 7620

gcgcaggggg aagtgccgcc cgcgctgcgg gaagtacggc tactggcgct ggacgttggc 7680

gctctgttgg ccggagccag catgaaaggc gagtttgaat cgcgtctgaa agggttactg 7740

gaagaggccg ggcgctcgcc gcagccggtt attctgtttg ttgatgaagt tcacactctg 7800

gtgggcgcgg gcggcgcatc cggcacgggc gatgccgcta acctgctgaa accggcgctg 7860

gcgcgcggca ccctgcggac tatcggcgcc accacctgga gcgaatacaa gcgccatatt 7920

gagaaagatc cggcgctgac ccgtcgtttt caggtgttgc agattgccga accggaagag 7980

atccccgcaa tggaaatggt gcgtggtctg gtggatacgc tggaaaaaca ccataacgta 8040

ctgattctgg atgaggcggt acgtgcggcg gtacagcttt ctcaccgcta cattcccgcc 8100

cggcagttgc cggataaggc catcagcctg ctggataccg ccgcggcccg cgtggcgctg 8160

acgctgcaca cgccgcctgc cagcgtacag ttcctgcgcc agcagctaaa agcggcggaa 8220

atggaacggt cgctgttgca gaagcaggaa aaaatgggga ttcagtcaga tgagcggcgc 8280

gatgcgctga tggcgcgaat tttctcgctc aacaatgaac tgactgcatc cgaatcccgc 8340

tggcagcggg agctggaact ggtacatacg ttgcaggaac tgcgtctcgc agagtctgat 8400

gctgatgaca aaaccacgct gcaacaggcc gaaacggcgc taagggagtg gcagggcgac 8460

gcgccggtgg tgttcccgga agtcagcgcg gcggttgtcg cggcgattgt cgccgactgg 8520

accggtattc ctgctgggcg catggtgaaa gatgaggcca gccaggtgct ggaactgcct 8580

gcccgactgg cgcaacgcgt taccgggcaa gacggcgcgc tggcgcagat tggtgaacgt 8640

attcagaccg ccagggcggg actgggcgat ccacgcaaac cggtgggcgt gtttatgctg 8700

gccgggccgt ccggtgtcgg taaaaccgaa accgcgctgg cgctggcgga ggctatctac 8760

ggcggcgagc agaacctggt aaccatcaat atgagcgagt tccaggaggc tcacaccgtt 8820

tccacgctga aaggcgcgcc gcccggctat gtgggctatg gcgagggtgg tgtgctgacg 8880

gaagcggtgc gtcgccaccc ctggagcgta gtgctgctcg acgagatcga aaaagcgcac 8940

catgacgtcc acgaactctt ctatcaggtg tttgacaagg gtgggatgga ggacggcgag 9000

ggtacacatg tcgatttcaa aaacaccacg ctattactca ccaccaacgt gggttccgac 9060

ctcatcagcc agatgtgtga agatccggcc ttaatgcccg atgctacggg gcttaaagag 9120

gcgctaatgc cggaattgcg caagcatttc ccggcggcat ttctgggccg cgtgacggtg 9180

atcccttacc tgccgctgga cgaaacgtcg cgtggcgtga ttgcccgtct gcaccttgac 9240

cggctggtgg cgcggatgag tgaacagcac ggcgtgacgc tgacgtatag cgaggaactg 9300

gtcgcacata ttgtggcgtg ctgtccaatg catgaaacgg gcgcgcggtt gctgattggc 9360

tacatcgaac agcacattct gccacgactg tcgcgctact ggttgcaggc catgacggaa 9420

aaagccgcga tcaggcagat tgatatcggc gttaatggtg atgagcagat tgtttttgag 9480

atcgtttgct gaaaccggcc gttcgaagtg tccgtagtgc gatttttaaaa actgtaccgg 9540

tataccgctc cccttgcggc aaccagttga ctaaaaagga aatgaaggat tatggctatc 9600

aacaatagcg cgcagaaatt catcgcgcgc aaccgcgcgc cgcgcgtgca gattgaatat 9660

gacgtagaga tttacggttc cgagaaaaaa atcgagctgc cgttcgtgat ggcggtgctg 9720

gccgatctgg ccgggaaacc gcgtgaagaa ctgccgccgg tgacggatcg caaattcctc 9780

gatattgata ttgataactt caatgagcgc atgaaagcca ttgcgccgcg cgtggcgttc 9840

gctgtgccga atacgctgac gggtgaaggt cagttgatgg tcgatatcac gctggaaaat 9900

atggacgact tttcgccggc acagattgcc cgcaaggtgg acgccctgaa ccagttactg 9960

gaagcccgca ctcaactggc gaacctccag acctacatgg atggcaaggc gggggcggaa 10020

aatctggtca ataaactgtt gcaggacccg actctgctga aaacgctggc gaatgcgccg 10080

aaatcggccg ctacccagca agatgtgtca gcggataatg aatcagcgga ataacgtcga 10140

atttttaagg aattttcatg gcaaacagta atatgcaggc aaccgacgcg gttgctcagg 10200

ataccgcctc cgcatccggt gaatttgatg cgttgctgaa tcaggccttc cgacccaaga 10260

ctacccaggc ggcaaaagcc gtggaagccg cggtgcagac gctggcgaac acgatcaccg 10320

tcagcgatga cgcctataaa agcatcagcg ctattattgc gcagatcgac tttaaactga 10380

ccgaacagat caaactgatc ctgcaacatc ccgactggca gaagctggaa tcctcgtggc 10440

gcggtatgga gcatctggtt tacaacaccg agaccgacga aaagctgaaa attcgcttca 10500

tgaatctgtc aaaagatgaa ttgcggcgca acatgaagcg ttacaagggc atcgcctggg 10560

atcaaagccc gatgttcaag aaactgtatg aagccgaata cggccagtta ggtggcgaac 10620

cttatggctg tatcattgcg gattactact tcgaccatac accgcccgat gtggatctgc 10680

ttggctctat cgccaaagtc gccgcgtcgg cccatgcgcc gtttattgcc ggggcttccc 10740

cctcggtact gcaaatggac tcctggcagg aactggcgaa tccccgcgac ctgaccaaaa 10800

tcgtcaccca gaacctggaa tatgcgccgt ggaactcgct gcgggctagc gaagactccc 10860

gttatattgg cctgacgatg ccgcgttttc ttgcccgcct gccgtatggc gcaaaaacca 10920

acccggtgga cgagtttgat tttgaagaag atgcggatgg ttctgaccat accaaatacg 10980

tctggagcaa cgcggcctac gcgatgggcg taaacatcaa ccgttccttc aaacactacg 11040

gctggtgtac gttgattcgc ggtgtggaat caggcggtgc ggtggaaaat cttccctgcc 11100

ataccttccc gactgacgat ggcggcgtgg acatgaaatg cccgaccgaa atcgccatct 11160

ctgaccgccg cgaggctgaa ctggcgaaaa acggttttat cccgttgatc caccgtaaaa 11220

actcagacta tgccgccttt atcggcgcac agtcgctgca aaaaccacag gaatactacg 11280

atccggacgc gacggccaac gctaacctgt ctgcccgtct accgtacctg ttcgcctgct 11340

cgcgcttcgc tcacttcctc aaatgtatcg tccgcgacaa aatcggttcc tttaaagagc 11400

gtgaggatat gcagcgctgg ctaaatgaat ggattatgaa ttatgtcgac gccgatccgg 11460

tgaactcctc gcaagaaact aaagcccgtc gtccgctggc tgccgctgaa gtagtggtgg 11520

aagaggtcga aggcaatcca ggttattacg acgcgaaatt cttcctgcgt ccgcatttcc 11580

agcttgaagg gctgacggga tcgctgcgcc tggtgacaaa actgccgtca gtgaagcagg 11640

gcaatgcctg atatatattt tgtgaatgtt taagcgagtg aagtcagaga agatagagaa 11700

tataaagagg gatatgaaga aaagaatttc gtctcgccca cggtctcgta aaggtggggt 11760

acgtaatgat gacacatatc cgaatgccag taacaatgcc gaagcttttt atatcattga 11820

gtaggaaata catattatga ccataagccc aacttttcat ctgttacctg gtattgttct 11880

gctcttttca caatatgctg tagcctggga agtcagttgc ccggctgtta ttgatactca 11940

gtcttctgct gtgagcctga agtctgatgt cccagcggcg tggcagcttt ctccccgata 12000

tatgtcgcgt ttatggttaa gtagtattgg ggtaacgcag ggtaaacctg aaaacctgat 12060

ggatctcaaa ccagagacta aaaaagtaaa cggtgaaaat tggtctgtat gggaaacaga 12120

acgtggtagc gataaagaaa ccgatcgcta ttgggtttcg tgtatttatg gtcatgaaca 12180

gatatggttg acgcaaccaa tacctgcttc ttctactcgc tgtaagactc gtaattttga 12240

gggatcgcca gaagaccagt ctgtatcttt tatctgtaat tagcgatttg agacgtgaaa 12300

atttcagtac aggttatggt ttttattatc ggaagttatg aagcattatt tatatgcatt 12360

aaataatgca aattcataaa ataactaaat acattatcgg taccggaaaa atatacagtc 12420

ctctgttctc ctgaagttat tggagaagga ttctgtacgg caatgattta tctataaaca 12480

aaaagatata gataaaatca ggtttatttt aagtaaaact taataaggat ataaaaatgg 12540

cttatgacat ttttttgaaa attgacggca ttgatggcga gtcaatggat gacaaacaca 12600

aaaatgaaat tgaagtactg agctggcgct ggaatattca tcaggaatcc accatgcacg 12660

ccggtagcgg cctcggctcc ggtaaggtct ccgtcaccaa cctggatttt gatcactata 12720

tcgaccgcgc cagcccgaac ctgttcaaat actgcgcctc cggcaagcac attccgcagg 12780

ccattctggt tatgcgtaag gctggcggca atccgctgga gtacctcaag tataccttca 12840

ccgacctgat tgtcgccgtg gtttccccga gcggcagcca cgatggtgaa atcgcctccc 12900

gtgaaacggt ggagctctcc ttcagcaccg tgaagcagga atacgtggtg cagaaccagc 12960

agggcggcag cggcggcacc atcaccgcag gctacgactt caaggccaac aaagaaattt 13020

aacggctgtt tttccggcca gatgttatgt ctggctggtt ttattgtttt gattttaaag 13080

gaatttacag tgaataaatg gcgtaacccc actgggtggt tatgtgcggt agctatgcct 13140

tttgcactgc tcctgctttc cggatgcggc agtagcgatt cgctacttga cccctaatcg 13200

cagcggcctg gcctgagcgt gaaagcgttt tacaaggtga attctgacaa tcagaagaaa 13260

gcggcgtcca tgaagatacg tgttgagaat taatgaccta cacagaattt ttagaggtta 13320

agcaaaatga acagaccttc attcaatgaa gcgtggttag cttttaggaa ggtgaatcat 13380

tccgtcgctg atgtgggtag cattattggt ggaaacgttg ggaaaaatat aactggtggt 13440

tattttcaaa atgcctgccc tattcgaatg agctatgttt tgaatgcgac agggttccca 13500

atagcccgta actctccgta tgcaaaggtt agtggtgccg ataataaatt ctatatttat 13560

cgcgtgaatg atatgattga ttatcttact catactatgg gcaagcctga tcttattgtt 13620

aataatccga aacagagtga ctttatcggg aagaaaggaa ttatcgtagt aaaagggcat 13680

ggctggagca atgccagagg acacgttaca ttatggaatg gcagtatctg ttcagatcag 13740

tgccacttat taaatgaccc agataatgga ccatttgttc ctgaagttgg gacactgtgg 13800

atactgccgt gaaatggtta atattagtta ctttcagcat aagtggaatg ctggtatggc 13860

agccatcctt tgcacaagaa gcattgacca cacaatattc acagtcggaa ctcttaaaaa 13920

attgggcgct gagtcattgt ctggcattag tatacaaaga tgatgtcgtt aaaaacgatg 13980

ccagagctac ggccagtgct taccttgaat atggtaaaca atctgtggag atttaccatg 14040

aaattgatga gattgcgaaa aaatattcag ggttgaaata taacggttcg atatcatcag 14100

attttaatac catgaagtgc atagatttta tccatgacag ggaattaaat gaattaatta 14160

aaaggcgtgt cgagaagtaa aattcaaaga tattaaaagc atacgttctc ttgctctgat 14220

gaatttatgg gtaagaaaga gctgtacagg aatagttaat ctgttcacct aataaagcag 14280

ataaatcagg gcttaattta ggtagttaaa aggatagtag atatgtctta tgacattttt 14340

ctgaaaattg acggcattga cggcgagtca atggatgaca aacacaaaaa tgaaattgaa 14400

gtactgagct ggcgctggaa tattcatcag gaatccacca tgcacgccgg tagcggtctc 14460

ggttccggta aggtctccgt cactaatctt tcatttgaac attacatcga tcgcgccagc 14520

ccgaacctgt tcaaatactg ctcttccggt aagcacattc cgcaggccat tctggttatg 14580

cgtaaggctg gcggcaatcc gctggagtac ctcaagtaca ctttcacaga tctgattatt 14640

gcaatggtat cgcccagcgg aagccaggga ggggaaattg cgtctcgcga atcaattgaa 14700

ctctccttca gcaccgtgaa gcaggaatac gtggtgcaga accagcaggg tggcagcggc 14760

ggcaccatca ccgcaggcta cgacttcaag gccaacaaag aaatttaacg gctgtttttc 14820

cggccagatt tatatctggc cggatttatt attttgattt taaaggaatt tacagtgaat 14880

gaatggcgta accccactcg gtggttatgt gcggtagcta tgccttttgc actgctcctg 14940

ctttccggat gcggcagtag cgatgcgtta cctgacctcg aatcacagcg actcgacctg 15000

agcgtgaaag cctccgataa ggtgaatcct gacaatcaga agaaggccgc gcccattgag 15060

atacgtgttt atgaactgaa aaatgacgcc gctttcacga cagctgatta ctggtcgctc 15120

catgacaacg acaaatccgt ccttaccgac gatttagtgc gtcgcgacag ctttattttg 15180

cgtcccggcg aagagaaaaa actgcgtcgc ccgctgaatg cgcagaccac ggcaatcggc 15240

gtactggccg gataccgtaa cctggccaaa tcggtctggc gggtaaccta caaaatcccg 15300

gaagccccgg aaaaagcctg gtacagcagc ttcatatcgg ggaaaggaaa agtgcagttg 15360

gaggcggaac tggaacaaag cgccattgta attacggaac gggataaatg aattatgagc 15420

tggaatgacc gcgtagtctg gagtgaagga caatttttac tgccgcagat gtttcagcag 15480

caagagcgtt atctggaaca cgtcatgcat taccgcagcc tgccgctgac ccccttttc 15540

tggggattca gccactacaa tattgatggc gaagcgctga acatcggtaa actgatactg 15600

aaagaggcat cagggatttt tcctgacggc acgccgttta acgcaccgga ccacaccccg 15660

ctgccgccgc cactgaccat tctgccggag cacctgaacc agcagatttg tctggcggta 15720

ccggtacgcg cgccgaacag cgaagaaacc acgtttgaca ataacccgga atcattggcg 15780

cgtttctcgg tacatgaaca cgacatccgc gacgccaact cgctgggacg tggcgcgcag 15840

ttattacagc tcagtcatct gcgcctgcgg ctgctgccgg aaaaggcggt gacgggcgcc 15900

tggattggcc tgccgttgac ccgcatcacc gggttgaacc ctgacgggcg gatagatatc 15960

gaccacgacc tgatcccgcc catcattaat tatcaggcca gttcactgat gtgtacctgg 16020

ctgtcgtgga tcaacgatct catccggatg cgggccgatt cgctggcgga acggctgacc 16080

ggcagcgaca accacggcca tgaagcagcg gaggtctccg attacctgct gctgcaaatt 16140

ctcaatcgct ttgagccgct gctgactcac ctggcgaaaa ccccgctggc cccggaggtg 16200

ctgtaccgct acctgtccga actggccggg gaactctcca cctatgtgcg tccacaaacg 16260

cgacggcccg ctgaatacaa agagtacaaa cacctgacgc cctatgccgg gttgaaatcg 16320

ctggttgatg aggtgcagtt cctgctgaac gcggtactga tccggggcgc gcagcgcatc 16380

gagctgaaag aggggactta cggcatcctg aatgcggtgg tggccccttc cgatcttgcc 16440

gatttcagca cgctggtact ggcgataaag gcttcaatgc cgaccgatgt gctactgcaa 16500

cattttgccg cccagaccaa aatcgggcca tccgatcgcc tgccggaact gatccgctcg 16560

catctgccgg ggctggcttt gcaggttctg cctgtaccac cgcgccaaat cccgtttcag 16620

gccggataca tctattacga catccgccgc gagggagcat tgtgggaaca cattgcccgt 16680

tacggcggga tggccatgca taccgccggg gaatttccgg ggctggagac agaactgtgg 16740

ggagtgcgcg ataaatgaca gacagtaccc tgacgccgcc agcggcggat atgatgtcct 16800

ttttgtccac cacgccggaa cataaggaca gtgaatatga aacgccggta cacaccagcc 16860

agcgcacgga actcaatgtc atcgttgaag acggtccgga cagcaaactc cggctggctg 16920

aaatcagcgc ggcggctaac ccgttgctcg ccgctgcccg gcctttattg tgcgctctcg 16980

cagccatgcc cgctaaactg gatgcggccc tggtagagcc ttaccgtaat ctgctggtac 17040

gcgagatgca tctgtaccag acattatgcg atcaggcgaa cctgcggcgc gagcacgtac 17100

tggcggtacg ttactgcctg tgtacggcgc ttgatgaagc cgccaataac acaacctggg 17160

gacggcgcgg cgtctgggcc ggaaaaagcc tgctggtaac atttcatggt gaaagcgaag 17220

gcgggataaa acttttccag atcatcgggc gtctggcggc cagcttccag gagcatggca 17280

acgtactgga ggttatctac cacctgctgg ggttgggatt tgaaggccgc tacagcgtgc 17340

agccagacgg gcgtaagcaa ctggacaata ttcgccagca actgctgaca cagctttcac 17400

agcgtcgcga tccggttatg cccgcgctct cgcctgactt tcaggggcg ataagcggac 17460

gactgcggcg gatgcgccgg gtgccggtct ggctgagcgc cgggatagcc ctgttggcga 17520

tgctgacgct gtttggcctt tacagccacc ggatggatgt gcagaccgtc accgtacaac 17580

agcatattga tgcgattggt ataaaactgc cgccgccgcc tgtgccggtt cataagctgc 17640

ggctgaaaat cctgctggca aacgaaatcg cccgtggcct gctgaccgtg gacgaagatg 17700

accagcacag tagggtggtc ttccgtggcg acgccatgtt tgtgccggga cagaaaacgg 17760

tgagtgacgc aatccggcca gtgattaaca aagcggcgcg ggaaatcgcc cgcgtgggcg 17820

gcgcagtcac tgtaacgggt cacactgaca gccagcccat tcattcggct gaattcccat 17880

ccaacctggt actgtcggaa aaacgggcgg cggaagttgc ggcgttgctg acctccggcg 17940

gcgtacctgc cggacgggta catatcgtcg gcaagggcga tacggtgccg gtggcggata 18000

acggcagtaa agccgggcgg gcgaaaaacc gtcgggtgga aattctggta gtggagtgaa 18060

tgaatgatga aaaaatcaac ctatgatgtg tctcatcatt cggcagtatg tggcgtgacg 18120

ggggattatt atcggatctc agcgacatat cacataacac gatctgttcg tgtttttttg 18180

atcatcttat gttgcctgtt atccggtggc gcttttgccg gatccccgat taacgcagga 18240

ttcatttccc ccgataatgt caacctcagt actcaggatt tcctgaaatt ttatgccact 18300

gacaacgtac agaaaaaaga caatgcactg atgtatatgc tgggggtgga ggatgcgaca 18360

gaaggtaaag cctggtgtgg atatggtcag gttgacagta taacaataaa ccatactgtg 18420

ctgacctggt ttgaacagca cgcagtgaaa aagcctgatg taagggcttc aatactaata 18480

gaggaagcat tagttaaaaa ttttccctgt cagaggacag actcctccat aaaaattgct 18540

tcccggtcat ctcccatttt atccctgacg ccggatgcgc ttaatctttc aggtaatgac 18600

ttttttaaat tttgggtgtc tggtaatcaa cgggataaac tcagggcggg tgtctatctg 18660

ctcggcgtgg aggatgcgac agagaacaaa ctgtggtgtg gatacgcttt atttaagacg 18720

ctaacattaa atgaattagt ctatgtttct cttaaaaata aaaccaatga ggaactgaat 18780

tctcgcgcgg ctgaacttat cataaataaa ttaatagagt atccctgtaa tatataaaat 18840

cattcaagtt gcatcaaggc ggcaagggag tgaatccccg ggagcgtaca ttagttcgtg 18900

actggggtga gcgaggaaag ccaacgcaca tgcagcctga agtatgacag gtataccctg 18960

tgataatgtt atcgctgcca gttcagataa agtcttggtg gtaaagtctt ggtggtaaga 19020

tattgatatg tttaaggtaa tggtctgtct gttggtgggt gttcctgcca tttcttatgc 19080

gcacgattat gggtgtgcta cggttggagc atcaatggag tcatcattat ttgatgccat 19140

aaaaaatgac ctgaatatag atgtcgctac tattataaag gataaaacaa aggtagagat 19200

tcttgatatc tcacctgtat ctaaggtcta tgcagaatct ctggccagga tggattatga 19260

aaaagataag gccaaaaata aagtggcaat attagataaa aaatcctatt tcgatagtta 19320

ctatgaaaat caggttaaat ctatcgtggc aaaatatacc tatattaaca aagataaaga 19380

aaaagatatt ttcattgcat ccagcttcat gaatgctgat gagtgttctg taagatttaa 19440

tggttatatt actttatcca gggaatttta aaatattgat tattgggtgg cgagtcgaaa 19500

tataatgcaa aaatttctta gtctgctttt ttcccggcgc gcgctggcag ttgtgggcgt 19560

tctggttctg gcgctgctgg tctggtttgt cgggccgctg gtgtcatttg ataccctgcg 19620

cccgctggcc tccgtgggta gccgggtagt gaccattgcc ctgttgctga tgctgctggt 19680

actgtggctg gtcaactggt cgatgagtat catcggcatc agtgtcctgt gcctggcgat 19740

tggcttcgtc acaccgctgc tggccctggg cgatgtccat ccgtttgcgc cgctgtgggt 19800

ccgcctgacc ctgattggtt tcatcctgct gatgtacgcg ctgtacggcc tgtaccggct 19860

gtggcgtgcg ctgcgtatgg atgaacaact gctgcgtcgc ttcctgcatc cgcgcgggga 19920

agaggtaccg gtggcaggcg agatcaaagc cgacctgcgc accgtcaacc atattgtcac 19980

gcaggccatc cggcagctgc ggcagttgcg ggtggatatg cctggctggc gtaaaatctt 20040

cgagggaaaa cgctttctgt atgagctgcc gtggttcatg gtggtcggca gtcccggcga 20100

cggcaaaacc acggccctgc tgaacaccgg attgcagttc ccgctggcgg agcaaatgga 20160

gcagacttcg cgcatcctga cagtaccggg tggcggcacg ctacactgcg actggtggtt 20220

taccaacgaa gcggtgttga ttgataccgc cggacgctac gcccgccacg atgacggtgg 20280

tgaagcgagc gccgcgcagc gtaacgccgg agagtggcag ggctttctcg gtctgctgcg 20340

taaacatcgc cccggcgcgc cgcttaacgg cgtgatcctg acgctaaacg tggcggattt 20400

aaccgcacag tcaccggcgg aacgcctggc ggcctgcgcc gctctgcggg cgcgactggc 20460

agaactgcgc gagaccctgg ggattcgctt tccggtctat ctggtggtca ccaaaatgga 20520

tttgttgccg gggttcagcg aatattttcg cacgctgacc agccatcttc gtgcacaaat 20580

ctggggcttc acgttgccgt acagccgcag gcgaaaagcg ggcgacccgc aggcgctgca 20640

cgccgcctgc gcgcaggagc tggcgcgcct gacgctgcgg ttggatcagg gactggatac 20700

ccggttacag gaagagtacg accttaaaag ccgccagcgg ctgtatacct tcccgcgtga 20760

gttcgccgcc ctcggcgagc cgttgctgga ggctattgaa cagatcttcc tcgattcaaa 20820

attcgatgcc acgcaactga ataacacgct gcgcggggtg tttttcacca gcgccgcgca 20880

ggcgcaggcc gacgccgtgg ccgaccagtt gagtatctgg cagcgctttg tccgggcgat 20940

aaaaaccgcc cgtggcgaat cctccgcctc tctcccacac gctctgccgg acggcaaccg 21000

cagctacttc ctgcatgacc tgctgacaca gtttatttttt cgtgaagcgc acctggtgga 21060

gccaaacctc cagtgggcct ggcgttaccg cctgctgcgc ctcggcgggc acctgctggt 21120

actggtgctg gcattcctgt tgtggcaggg gatgcagacc agccagcaga ccaacggcga 21180

ctatctgaat gaaatcagcg cccgcgcgac ccggctggac ggtgatgtga aagcctacac 21240

cggtaaaccg gcgatggctc ccgtcccggc actgctggac agcgcaaggg aactgtccgc 21300

ctggccggaa ctggacccgg acgcgccgcc gctggcctgg cgctacggtc tgtacagcgt 21360

accgccggta accgacagcg tggcgtcgct gtacaaccgt ctgctggatc aactgctgct 21420

gccgccgctg gtgaaacgga tggagtatgt gctggcggac gccattgccc gtcaggatag 21480

taaagcggcc tacgatgccc tgcgcatcta tctgctactg aatctggata aagatcacga 21540

agataaatac aacgcggcgg agatccagtc gtgggtgatt aacgatctgg ggaacagcga 21600

cagcgtggcc gggttcggcg ggcgcgccgc cgtgctgacg catatcgaag cgctgtttga 21660

cggcagccgg gtggtgcatt caccgtatga gaaagatgag gcgctgatcc gccaggcgcg 21720

ggcattcctc gacggtcaca ccagtaccga gcgtatctac gcgcgggcgc tggcggcaat 21780

ggagagcgaa gcgccgcagg agttcacgct ggtacgcgcc gtcggcgcgg atgcgggaac 21840

ggtctttgtg cgtagcaacg gcgcgccgct ggatcggggc gtgccgggta tttttacccg 21900

tgaaggatac cgggagctgt tcgacaaacg attaccggaa tttgtggcgg cggcgacggc 21960

gaacgatggc tgggtgatgg gccgggagag tacgccaaaa aagctgactg acagcctgcg 22020

cagccagata ccggggcagg agcagtctgt cgcccgcgaa gtccgccgtt tgtacctgac 22080

ggaatatgcc cgccgctggc aggattttct ggacagtatc catagtatca acagtgccgg 22140

ggaagagggc agttccggcc tggcctatga tttacaggtg ctgcgcaccc tggcgtcgcc 22200

ggactcaccg ctgatgcggc tgggaaaagc ggtggtggag cagaccacgc tggtgccgcc 22260

gccggacccg caggccagac agaaacaact ggcgcagcgc gcatccggta acgcggggaa 22320

agtggtacag acggcaaaac tgttccagga tattcacccg gaagaacggc tggaaaaaac 22380

gctggtggat gatcgtttcg ccgcgctgcg cgaggtcatt gccgggcgga cggacggcgg 22440

acagagcggc ggtgggacga tgcagatcgc ttcgctgctg accatgctca acgagtatta 22500

cacccaactg accattgctg atagcgcgct ggcggcgggg acgttgccag cgcgcattac 22560

tgccgcagac aagttgcaac tggaggcggc gaaactgccc gcgccgctga aaaacatcct 22620

gctggatctg acgaagcagg gaacgcgcaa aatcaacgcc gggaccggcg acgtgctgaa 22680

cacccagatg gaggcgatga tgggcgacga ctgccgtgac gccatcgacg ggcgctatcc 22740

gttcgccgac agtccgcagg aggtcagcgc cgaggacttt aaccgcatct tcgccagcgg 22800

cggcgtactg gatgctttct ggagcaaaca actggctccg ctggcggaca ccgccagcga 22860

cccgtggcgc tacaaaccga ccgaaggcaa catgacattg caggggccgg atctgacgcc 22920

gtttcagcag gcgaagcaga tccgcagcgt atttttaac agcgagggcg ggaaaaaatt 22980

ttcctggtcg atgcagatta gcgtggtgga tatggacccg gccatcacgg aactggtgat 23040

tgatatcgac ggccaggtgc tgcgctacgc ccatggtccg gaccgtcccc tgaaagtaac 23100

gtggccggga ccgcgcaacg gctcgatggc ggaaatcacc gccagcccgc gtattcgcca 23160

ggacacctca acgctactga ccggcgggcc gtgggcgctg tttcatctgc tggacgccgg 23220

aatggtacag gaaacggcgg tgcgggggcg tcaactggtg gaatatgatt ttgacggtcg 23280

ccgggtggtg ctggaaatca ccgccgggcg ggattttaac ccggtcagcc gggagctgtt 23340

gcagaacttc agttgcccgg cgagggcgct gtaatgcgcc ggccagcagt ggtatcagcc 23400

ccggcgatgt tcggcaggct gccggaccag cgcgactatg tgcgctggcg agtgggcgct 23460

gaagagggac gaatatggca ggactggctg aaccgccaga cgtgggtggg cagcgggcgg 23520

catatcgtgt tgccgggagg ccagacgacc gatgaagagc gtgacggctg gatgcacctg 23580

tcgccgcgtg gcgaacatga tatgcccgcg ccggaacacc atccgctgcc gtggagcttt 23640

gtgatgtcgc cgggattttt gccgatgggt gggacgactg actggctgac cggagtgctg 23700

atggcgtcac gggatagcgt gggacgcccg tggccgctgg tgatttacca gcgctgcggt 23760

cgggagtggc tggacgagag tttgcaggag acgcagggct ggctgtactg gctggcgagg 23820

ctggcggcgc agcatatcac gccggacacg atgcggcgcg ggcggctgac cgagcaggtg 23880

gaccagctat gggcgatgtg gcagccgggg ccgtggtggg cgcagtggct gcgtgggtta 23940

cgacgtacat cgcaacgcag ccgggaactg accgggctgc cggatgaagc accggtagta 24000

gaactgccgg gcgtgcgtta tctgccgtgg ccgggctggc cgggtaagac gctggggcag 24060

gcaacaccgg ggcaaggctg gttctggcag cagaacagtg aagggcggta tgtggatgcg 24120

ctgcggttgg tggatgaaaa aaatagaggt cagatgctgt gaaaaagtac gctgtagagg 24180

ttctgtttat gtctgcctgt gcaggaatgt tcctgcccgt atttgcctgg ggcggaacag 24240

atgtaaacat tgataacccg ctggcggagt gtgtggatat ccatcccgtt catcgtcagg 24300

aaatggataa ccttaccatt ctgaaaacaa cggtcacact gaaaaaaagc accggtgaat 24360

gtggctgttt ctcaacatta atcaactata ccagcttact ggcgcaggat gttgagggtt 24420

acggacgcgg aagtgcatat tcccttcagg aggggaatat ctcactggcg aagatgcagg 24480

gacgttatcc tttcagcttt gtattatcgg tggataacca gtctgtacgg gatcagaagc 24540

tggcgctaat gatacgttgt acaccaccgc tgtaatacac agaatagtca gggagaagat 24600

gatggcagta agactgactt ttgacgggca aaagctgaca tggcctggta tcgggatatt 24660

taaggcgacc acggggttac cggatttaca gtggccagat aaacagtgtg tgccggatgc 24720

ggcgataccg gaagggaatt ataaattgtt tattcagttt caggggggagg caccgataag 24780

aaatgctgcg gattgtgatc tgggaccatc atggggctgg agtaccattc cgcgaggcca 24840

ggctgccgga acatgtgaga tatactgggc gaactgggga tataatcgta tccggctgga 24900

atcagcggat gagaagaccc gaaaagcctg tgggggcaag cggggtggtt tttatatcca 24960

tgattccacc aaaggttaca gtcatggttg tattgaagtg gaaccggtgt ttttccgtat 25020

tctgaaacag gagacggaaa aagaaaatgg tgaaaagaca tttacggtta atgttaagta 25080

tgtttctggt cagcaaacga atggtggaac aaaacaataa taccgttaaa accggatgaa 25140

attatcgtgt ctggacgtgc tgtggttgac aatgattaca atggtggcca gataagtaat 25200

gagcaccaat gataaagtat atgactggcg gtgtgctggt tattaatcaa atttctataa 25260

aatgcaatgc gaatctggta agcgtaataa aaataaaatc ataatattgt tatgttcatt 25320

tccttattta tgtaattcag tatttatgtt atgtgctaat ttttgtgttt ttatttttcat 25380

ggctcttgtc agcaatatac cctgttcttc tggtaaatat ttaaatataa caggctggtt 25440

gcattataaa gtgcggggca ctgtttcctg acggtgagtc tatttatttt aatccggtat 25500

taaaggagtc actaccatga gttttgtatc cacaaataat aaatccggta tgggagggct 25560

gacgacaacc acgccgccga taaccggaga aagtggcggt gtcaccgcag attcagtcgc 25620

cggaagcgtg gcagatgcgg cggaatccgc cgtggaacag gctgcgggat cgctatttgg 25680

cgcattgccg gagccatcag gactggtgaa agccgcggta gcagcggcgc aggctgccgc 25740

cgccgcaggt atggcgcagg atgcggtatc ggccatcgtc tctgctgttg caggcgggcc 25800

gggggcgcat aatgtgacgg tcagcggcag cgccgtaccg ccgggcgcat tactgttcgc 25860

cagcctggac ggcggcgaaa cattaagtga actgttcagc tatgtggtac agctaaaaac 25920

gcccgacacc ctgaatctgg gctatgtctc cccggcggcc aacctgccgc tcaaaccgat 25980

ggtgggcaaa gatctgtgcg tcaacatcga actggatggt ggcggtaaac gacatatcag 26040

cgggctggtc acggcggcgc gggtggtggg ccatgaaggg cgttcggtta cctatgagct 26100

gcgtatggag ccgtgggtaa aactgctgac ccataccagc gactacaaag cattccagaa 26160

taaaaccgtg gtggatattc tggatgaggt tctggcggaa tatccctacc cggtggaaaa 26220

gcggctggtg gaaagctacc cggtacgcac ctggcaggtg cagtacggtg aaactgattt 26280

tgattttctt cagcgactga tgcaggagtg gggcatctac tggtggtttg agcacagcga 26340

ggacagccac acgctggtgc tggcggatgc catcagcgcc cacaaagcat gtccggactc 26400

gccgctggtc gagtggcacc aggaagggct gaagctggac aaggagttta tccacactat 26460

cacggcaaac gagagcctgc ggactggaca gtgggtgctg gatgatttcg attttacgaa 26520

gccacgttca ttgctggcaa acaccgtggc aaacccgcgt gaaaccggtc atgccaccta 26580

cgagcattat gagtggccgg gagactactt cgacaagagt gaaggcgaga tgctgacgcg 26640

cattcgtatg gaagcgcagc gcagccccgg cagtcgggtg ctggggggag ggaatatccg 26700

cacactcatg accggttata ccttcacgct ggaaaactat cccaccgccg aagtcaatca 26760

ggaatatctg ctgatgcaga ccttgctgtt tgtgcaggac aacgcgcagc acagcgggca 26820

ggaccagcac tttacctttt ccacccgttt tgaactgcac cccacccgcg aggtgttccg 26880

cccgcagcgg acggtgagca aaccccacac caaagggccg cagagcgcca tcgtcaccgg 26940

cccggcgggc caggaaatct ggacggatca gtacgggcgg gtaaaggtac agtttggttg 27000

ggatcgctac ggcaaaatgg atgaaaacag cacctgctgg atacgcgtca gctacccgtg 27060

ggcgggcaaa ggcttcggga tgatccagat cccgcgtatc ggccaggaag tgctggtgga 27120

tttcaaaaac ggcgatccgg atctgccgat catcgtgggg cgtacctaca accaggacac 27180

catgccgccg tggggactgc cgggaatggc gtcgcagagc gggatcttca gccactcgct 27240

gtatggcggg ccaacgaacg gcaacatgct gcgttttgac gacaaaacgg gcgcggagga 27300

agtgaagttc cacgcggaaa aagatctcaa caccacggtg aagaataatg aaacgcatac 27360

ggttatggtg gatcgcacta aaaccattat taaaaatgaa accaacagta ttggtgagga 27420

cagaaacacc acggtaacga agaatgacgg cctttccgta aaactggcgc agacgatcaa 27480

tatcggcacc acttatcgtt tagatgttgg cgatcaattc acgcttcgct gcggcaatgc 27540

ggcgcttgtt ttacataagg acggctccat tgagttttgt ggcaagcaac tgatgttaca 27600

taccagcgat gtcatgcaac tgattggtaa aggtattgat atgaacccgg atggcggcac 27660

agccgtaacc gccgatgata ttgcccccct tctcacctct gagtgatctg aattaaacct 27720

ggagttctca tggatcgacc ataccgcata caggaagggt gttttgtcct gcctgaaaca 27780

tttacggatc gcagcgtcaa tatttttatc ctggagggca atgaacgaac atcgcccagc 27840

ctgaatattt cccgcgatac gctaaaacct gatgaagacc tgcccgccta tattgaccgc 27900

cagattgcac tgatgaaaaa aaatctcggt cagcaccggg tattgtcgcg agcgcctgca 27960

caggcaggaa cgggcaatga tgcccttatg ggggaacaaa ttgccgccac ccataaatcc 28020

gggaaaacgg aagtgtacca gcgtcaggcc gggtttattg caacccctgg caaggtactg 28080

gtcttcaccc tgaccagtcc ccgtcctttt gatgataaag cagacctact ctggaacacc 28140

tggctggcag gctttcagcc ggataaaaac gaataatcac acggaggtgt gaccatgtat 28200

gaagcagccc gtgtggatga tcctatctac cacaccagcg cgctcgccgg gtttcttatc 28260

ggcgctatca tcggcatcgc cattatcgcg cttgccgcct ttgccttctt tagctgcggt 28320

tttcttgccg ggctgattct gggttttatg gccgatcaaa tagcctccgg ggtattgcaa 28380

ctgggcgagg ccatcgggcg ctccatccac cacacggcag gaaaaatcct caccggttcg 28440

gagaatgtca gcaccaacag tcgcccggcg gcgcgcgcgg tactgagtac ggtgaaatgc 28500

gataaccata tcgcagaaaa acgcatcgcc caagggtcgg aaaatatcta catcaacagc 28560

cagcccgccg cccgtaagga tgaccacacc gaatgcgacg cggtgattga agacggttcg 28620

ccgaatgtgt ttctcggcgg cggcacacag acggtactgg aaatcagttc tgaaattccg 28680

gactggctgc gcaaggtggt ggatgtattg tttgtcgtgg cgagtctgct cggcgggctg 28740

gccggggcgt ggcggcaggc ggcaaagctg gggacgaaat ttggcactaa atgtgccgct 28800

aagtttatcg gcggggagct tgtcgggatg gccgtgggtg aggctatcag cgggctgttc 28860

agcaatccgg tggatgtgac caccgggcag aaaatcctgc tgccggaaac ggacttcacc 28920

ctgcccggtc gcctgccggt cacctgctcg cgtttttacg ccagccacct ggaaactgtg 28980

ggactgttgg gacggggctg gcggctgaac tgggaaacca gcctgcgcga tgacgatgaa 29040

cacatcacgc tgaccggcgt acaggggcgg gaactgcgtt acccgaaaac gatgctgacg 29100

cccggccacc agatatttga cccggaagaa cagttatacc tcagccgcct gcatgacggg 29160

cgttacgtgc tgcattacac cgatcgcagc tattacgtat ttggtgattt tgacagtgac 29220

ggcatggcat acctgctgtt tatggagacg ccgcaccgcc agcgcattgt cttcgggcac 29280

gaaggaggca gactggtacg gatagcctcc agcagcgggc atcacctgtt actgcaccgc 29340

acacagaccc cggcagggga gcggctgtcg cgaattgaac tggtgcaggg cggcacccgt 29400

ggcaatctgg tggagtaccg gtatgacgat aacggtcaac tgaccggcgt ggtgaaccgg 29460

gcgggaacgc aggtgcgtca gtttgcttat gaaaacgggc tgatgacggc gcacagcaat 29520

gcgacggggt tcacctgccg ctaccgctgg caggaactcg acggcgcgcc gcgcgtgacg 29580

gagcacgaca ccagtgacgg cgaacattac cgctttgact atgattttgc cgcaggcacc 29640

accaccgtca ccggcaggca gggggagaca tggcagtggt ggtacgacag ggaaacgtat 29700

atcaccgcgc accggacgcc gggcggtgga atgtaccgct tcacgtacaa cgaagaccac 29760

ttccctgtca acattgagct gcccggcggt cgcacggtgg cgtatgaata tgacatccag 29820

aaccgggtgg tgaagacgac agatccggaa ggccgggtga cgcagacgca gtggaacggc 29880

gagttcgacg aaatcacgcg cacggcgctg gacgatgacg ctgtctggaa aacgcagtac 29940

aacgcccacg gccagccagt gcaggagacg gacccggaag ggcgggtgac gcagtacgct 30000

tacgatgaac aggggcagat gtgcagccgg acggatgcgg cgggcggcac gcagggcggc 30060

gtgcggcggg agacgcagca gcgggatgcg ctgggccgtc tgttacggac ggagaatgaa 30120

cacggccagc ggacgttcag ctacaaccgg ctggaccaga taacggcagt gacgctcacg 30180

cccacggagg cggggcaaca gcagcaccgg atgcaggccg acacggtgcg ttttgagtat 30240

gaccgcagcg gctggctgac ggcggagcac gcggggaacg gtagcatatg ttatcagcgc 30300

gacgcgctgg gcaacccgac ggacatcacg ctgccggacg ggcagcacct gacgcatctg 30360

tattacggga gcgggcatct gttacagacg gcgctggacg gcctgacggt gagcgagtat 30420

gagcgcgaca gcctgcaccg tcagataatg cgcacgcagg ggcagcttgc gacgtacagc 30480

ggctatgacg acgacgggct gctgagctgg cagcgcagtc tggcgtccgg cagtgcccct 30540

gttcttcctg gccagcgccc ggcgcggcag ggctgcgtga cgtcgaggga ctattactgg 30600

aacaaccacg gcgaggtggg cacgattgac gacggcctgc gtggcagcgt ggtgtacagc 30660

tatgacagaa gcggttacct gaccgggcgc tcaggtcaga tgtatgacca tgaccgttat 30720

tattacgata aggcgggcaa cctgctggat aacgaagggc agggagcggt gatgagcaac 30780

cggctgccgg gctgtggtcg tgaccgttac ggctataacg agtggggcga gctgaccacg 30840

cggcgcgacc agcaactgga gtggaacgcg caggggcagc tgacgcgggt catcagcggc 30900

aacacggaga cgcactacgg ctacgatgcg ctggggaggc gaacccgcaa ggcgacgtac 30960

gggcggcaca cgggccatac ggcgcggagc cggacggact ttgtgtggga ggggttcagg 31020

ctgttgcagg agaacgtgca gcagcagggc tggcggacct atctgtacga tgcggaacag 31080

ccgtacacgc cggtggcgag cgtgacgggg cggggagaaa gcaggcaggt gtggtattac 31140

cacacggatg tgacgggcac gccgcaggag gtgacggcgg cggacggaac gctggtgtgg 31200

gcggggtata tcagggggtt tggagagaat gcggcggaca tcagcaacag cggggcgtac 31260

tttcaccagc cgctgcggct gccggggcag tattttgacg acgagacagg gctgcattac 31320

aatctgttca gatattatgc accggagtgt ggacggtttg tcagtcagga tccgatcggg 31380

ctgaggggcg ggttaaacct ttatcagtat gcgccaaatc ctctcaaata tatagaccca 31440

cttggtttaa ccgcgactgt tgggcgatgg atggggcctg cggaatatca gcaaatgctt 31500

gatactggga cagtagtaca aagttcaaca gggacaactc atgttgccta ccctgctgat 31560

atagatgctt ttggtaagca agcaaaaaat ggtgctatgt atgttgaatt tgatgtgcct 31620

gaaaaatcat tagtacctac aaatgaagga tgggcaaaaa tagtagggcc agattctatc 31680

gaagggcgat tagctaaacg caaaggtttg cctgttcctg aaatgccaac agcagaaaac 31740

ataactgtaa ggggcgagaa aattaatggg gaagttgaag caaaatgcta aataaattta 31800

aattgtgggt gagcaaacat actgattata cggtaattca taatgaaaat gatttatctt 31860

acagtattat tatagatttt gaagatgacc ggtatatatc aagatttact gtatgggatg 31920

acctaagctg tatgtcagaa gtaatggatg tggatactgg tttatataaa ttaaacaaga 31980

gaaacgaatt ttctacattt gatgaacttc tggatatatt tgatgatttt atgataagta 32040

ttaaataata gttggccggg taagaagtta actcttcccg gctgttttat tatctaaccc 32100

ccatcaatcc ggagacgcgc taccggtacg atgcgctggg caggcgggtg agcaaggcga 32160

cgtacgggcg gcacacgggc catacggcgc ggagccggac ggactttgtg tgggaggggt 32220

tcaggctgtt gcaggagaac gtgcagcagc agggctggcg gacatatctg tacgatgcgg 32280

aacagccgta cacgccggtg gcgagcgtga cgggaaaggg agaaagcagg caggtgtggt 32340

attaccacac ggacgtgacg ggcacgccgc aggaggtgac ggcggcggac ggaacgctgg 32400

tgtgggcggg gtatatcagg gggtttggag aaaatgcggc ggacatcagc aacagcgggg 32460

cgtactttca ccagccgctg cggctgccgg ggcagtattt tgacgacgag acagggctgc 32520

attacaatct gttcagatat tatgcaccgg agtgtggacg gttcgtcagt caggatccga 32580

ttgggctggc gggggggget gaatctttac cagtatgcgc ctaatccgat tagatggatc 32640

gatcctttag gacttgctat cctggagcat caatctaatt ttgatgcggc aaggagaacc 32700

ggatttgaaa atgcgggtat gacaaaccct gaggatgtca ctttctcgaa agtcgatccc 32760

aaaactggta ctgttgttga gtttaaaggt ccaaatgggg ctaaagttgc ttatgatgca 32820

cctcatgcag atatggatgt gacagcaggg catgataaac cacatgttgg ttggcaatcc 32880

gcaggaaaaa gaggttccgg aggagctaat agaggtaata ttacttatga tggcccacaa 32940

catccgcatc gctctgactc taagggagat gataaatgtt aaattcaaat atgtctgaac 33000

ttagaatcga actggagaat gcgattaaaa atctcggtat tcatgattat cgtgtcgata 33060

aacccgaaca aatcgtttct gagataaaag agatatatgt taatggtaat cctagaacct 33120

ggtggttatc attaaaacat agacaatatg tcttttctta taccgataat tctggatata 33180

aaaacatatc acaaatagta agtaaacaac tcaatgaaag caatgtaatc aacaaacata 33240

tatttttgat tgctgatgaa gataatgagc aaatatatgt atataacgtt cctcttaact 33300

ccctgcctga aattatagaa aattgcagat attttgaata ttatgttgca gatcatgaac 33360

tatcttggct tatatgtgaa aatgatcatg gtgatttgat tgtatgctca accattaagt 33420

aaagcgcgag tgctctttag cgatatagtt gcccatattt aggcgttact agccgaagat 33480

ggcgcgattg tctggcaggg gaaacagcaa ttctgaggtc aggaagatag cataacccat 33540

taaccgggat agatccgcta gacctgaatc cagttgatgc gacaggttat agggtttatg 33600

gttatttgct cctggagcaa ataaacctta ttacattggt attactaatg atatggtttg 33660

acgaagggcc gagcattaaa gcactggcag gttatcaaaa gaaaatggaa ggatgctgcc 33720

atttgatgaa aatgtaatct aatggaaagt cagaggttac gagaaatatt atatagagaa 33780

atataaaacc agaaccggaa ccataggtga aaaaattccc tcaacaaata gagagaataa 33840

atataattca tttgatcatg ggcgaacaga tcccagcgca caagcattta aagactctta 33900

aaatagtaag ggagttggtt ccggtggagg aaaatgcgga tgagtgatta agaattttgg 33960

ggctgtgata agaagtcgag aacaatgctg cgttttgtga agcccggaga catattttgt 34020

tttaaattag atgaagatag atattgtttt gggcgaatta taacactaat gactgtcgga 34080

catctttctg aattatttga tataattaaa aaacccccctg gaataacaga gttagaaatt 34140

agtaatgcaa ggcgaattat tgaaccaatt atagtggata catattcttt atttgataag 34200

aaattagaaa atggaagtga ctggagaatt attggtcatc aggttaatta caatccaaaa 34260

aatttagatg gtatctattt tgcacttgga ataggtgatt cctgtaaaaa gaaagactgt 34320

tacggaaatg attttctcat ttcagaaagt gagtggaaaa cacttcctaa attatctcct 34380

aaaggggttt ttgatatcaa aaaacggctt gaaattgcct gaaaatgaaa ataaaaagcc 34440

gggaaagatc ttttgtcttc ccggatttta ttatttaatc cccgttcacc acattattta 34500

cccccgcctt aatatgcttc atcgactttt tcacctgata aagctccttc cgtagatccc 34560

tcacttcgtc cgtctctgca atcaggatca aacacccctc ggagatcttc acggtgacgc 34620

cgtgcccggt ctcaaatccc gcttcttcct gccagtcacc cttaaggtgc tggctgggga 34680

tttgtgagta acaggcggtc atgcaggttt cgctgttgat atggcggacg ctgacgccgg 34740

aggcattcat atttgctgac taaataaatt cttatttatc cgccggatgc tggctgattg 34800

tggagctcag ggtgagtgag tatgggcgcg acatcctggc accgcgcctc cctctccccg 34860

gccagccccg gccggtgatg agcagccggc tgccggggcc gtattttgac gatgaaacgg 34920

gcctgcatta aaatctgttc agatattatg taccggagtg tggctggttc gtcagtcagg 34980

atccaatagg gctgaaaggg ggatggaacc gatatcattc tccgctgaat cctattacag 35040

atagtgatcc tcttggcctt attacttgtg gtgctgatag aggtgattct ggcaagttat 35100

taagatgaaa aatggttgaa aataattgcg ttagtgaatg ctcatttatg cctttaccta 35160

caaaaagtaa cggatttgct tgctggaatt gtgttaacga atgcaaataa aatcaatgct 35220

gttgatgttt ttactacagg aataagtatg gttaacgata aggatacagc tatattaatt 35280

agtaatttaa tgttgagatt cggtaaggag cttgatgaat ctgttgctgt tgttcagtcc 35340

cgttgtgatg aggatgaatt taatgtatat cgagaaacgg ttggttttat catgggtgaa 35400

atgcttatta aaataatgaa tccattatat gaaaaacatc cagaaataaa accaaaagga 35460

ttgaaacaaa acatctggaa ccggatgaat aatgtgtaaa agccggaggg gttatctttt 35520

cccggctttt tattatcaat tactcattaa ctcctgttcc gttcttttgc gtttaatcac 35580

cggaatatct ccggtattgt tcagcgcccc ggaaatgttt ttaaccactg ttctgcactc 35640

cgtttattaa tgcgggttac gcccatccct tcaatacagc caaagagtcc gtgggtatgc 35700

tgcggcgtga tcacgatgca atccctcatt acccgcacct tgacgggcat cccgttaata 35760

aaccccgcct gcggcatcca ttcaccgtac aaacgaactg aaagtctctc aacgcgtgag 35820

tatgtaagta tcccgcataa tcgagccatt cacatttaga gatcatccga cataatcaat 35880

ctgccaacgc aggagatcgc tatgcgtaaa gcccgtatta ctgcgcacca gatcatcgct 35940

gtgattagat cagtcgaatc cggacggact gttaaagatg tctaccggga gcccggtatt 36000

tctgaagcca ccagggacaa ctggaggtct ggatacggcg gcagggatac gcgtggaatc 36060

acaaaaggct ccaccgtatt tactgtctgc tcaagctgaa ttttcgccgt aagggtaaac 36120

aacggctgcc ggtacgcaat ccctcgccac tggtcacgcc ggaagcgctg aaccagagct 36180

ggtctgtggg cgtcgttttc gcacgttcaa tgttgttgat gactgtaatc gtgaagcgtt 36240

gtcgattgaa atcgatctga atctgccagc tctgcgagtg gtccgtgtac tcgacaggat 36300

tacagcaacc gcggttatct ggccatgctg cgtatggata agggaccgga atttatctcg 36360

ctggcactgg ctgaatgggc aaagaaacat gcagtaaagc tggcgtttat ccagccgggt 36420

aagccgaaga aaaacgtttt catcacgcgc tttaaccgga cataccgtac agaaatactc 36480

aattcttatc tgttcagaac gctgaatgag gtgtgggaaa ttacggataa agggttatca 36540

gaatataact gcgaacgtcc acatgaatcg cggaacaata tgataccgaa ggaataccgc 36600

caataacgtt atctggccgg aatcttaaaa atgcatggaa ctaaaacggg tctatttaca 36660

ggggcacctg cgatgaattt cgctgcactg aaaagcgata ccggatgaga gctgcttcaa 36720

attaatgtgc catgttcacg gggaggttgt gcgacgtttg cataatccag caagaactga 36780

aaggaagggg agagcttttt catgcctgta taatcagtct ggcctgtgtc agtcagctct 36840

tagtgttgag actctcgttg gagcgttata attgcttttc tgtttcggaa aacaagattt 36900

tccattaaag atcttccctg cgaggaaaag ttaactaata atcttaccgt cgagttagga 36960

gatgtatgtt taaatataaa caatgttgca acgatgcctg ataattatcc tctcttcgaa 37020

gataagtttc ccacacccag tgtagtaggt gtcatggtaa tgttatcact tgaatgtaaa 37080

tggaaggtat aattgctttt tgactggcat tctattccac cctgacaaca cgatgttaac 37140

atcaacactg tttatattgg caataacgca atttttttca gattaagagg tgctctgata 37200

```
tatagatttt tatgacatta cttatttgaa ttggtaacaa ataccaataa gtacaagctg 37260

ttattaccag ccacggattt tttacatacg gtaagatttg gtatggcgtt atgtattctg 37320

gatgtgctgg attatttttaa tttggtttaa aaaaggtggt tattcaa atg aaa agc   37376
                                                    Met Lys Ser
                                                      1

ata aaa aaa ttg att atc gca agt gcg ttg agc atg atg gct gct agt   37424
Ile Lys Lys Leu Ile Ile Ala Ser Ala Leu Ser Met Met Ala Ala Ser
  5                  10                  15

tgt tat gct ggc tca ttt ttg ccg aac tca gag caa caa aaa tca gtg   37472
Cys Tyr Ala Gly Ser Phe Leu Pro Asn Ser Glu Gln Gln Lys Ser Val
 20                  25                  30                  35

gat att gtg ttt tcc tct ccc caa gat tta acc gta tcg ctt att cca   37520
Asp Ile Val Phe Ser Ser Pro Gln Asp Leu Thr Val Ser Leu Ile Pro
                 40                  45                  50

gtg tcg ggc tta aag gct ggg aaa aat gct cct agc gcg aaa att gcg   37568
Val Ser Gly Leu Lys Ala Gly Lys Asn Ala Pro Ser Ala Lys Ile Ala
             55                  60                  65

aag ctt gta gtt aat tct act act ctt aaa gaa ttc ggg gtc agg ggg   37616
Lys Leu Val Val Asn Ser Thr Thr Leu Lys Glu Phe Gly Val Arg Gly
         70                  75                  80

att tct aac aac gtg gta gac agt act ggc act gca tgg cgt gta gct   37664
Ile Ser Asn Asn Val Val Asp Ser Thr Gly Thr Ala Trp Arg Val Ala
     85                  90                  95

ggt aaa aat act ggt aaa gag atc ggt gtg ggc tta tca agt gac agt   37712
Gly Lys Asn Thr Gly Lys Glu Ile Gly Val Gly Leu Ser Ser Asp Ser
100                 105                 110                 115

ctt aga aga tct gat agc acg gaa aaa tgg aat ggg gtg aac tgg atg   37760
Leu Arg Arg Ser Asp Ser Thr Glu Lys Trp Asn Gly Val Asn Trp Met
                120                 125                 130

acc ttt aat agc aat gac aca ctt gat att gtc ctg aca gga ccg gcg   37808
Thr Phe Asn Ser Asn Asp Thr Leu Asp Ile Val Leu Thr Gly Pro Ala
            135                 140                 145

cag aat gtc aca gct gac acg tac cca ata act tta gac gta gtg gga   37856
Gln Asn Val Thr Ala Asp Thr Tyr Pro Ile Thr Leu Asp Val Val Gly
        150                 155                 160

tat caa cct taa tagtaaacaa ctattagtgt attgtgcctt gtttaaggcg       37908
Tyr Gln Pro
    165

caatacacat caaatcatct atttttcttt tacaattttt gat atg aaa ata gtt   37963
                                              Met Lys Ile Val
                                                  170

aat ttt gct gta atg gcg gta gct ttg ttc gcc act aat tct atg gtt   38011
Asn Phe Ala Val Met Ala Val Ala Leu Phe Ala Thr Asn Ser Met Val
            175                 180                 185
```

```
tca gta tat gcc gtc aac cag caa tta aat tca gcc act aaa tta ttc   38059
Ser Val Tyr Ala Val Asn Gln Gln Leu Asn Ser Ala Thr Lys Leu Phe
        190                 195                 200

agc gtg aag ctg ggg gct aca cga gtg att tat cac gct ggt acg gct   38107
Ser Val Lys Leu Gly Ala Thr Arg Val Ile Tyr His Ala Gly Thr Ala
    205                 210                 215

gga gcg acg ctc tcg gtg agc aac ccg cag aat tac cct att ttg gtt   38155
Gly Ala Thr Leu Ser Val Ser Asn Pro Gln Asn Tyr Pro Ile Leu Val
220                 225                 230                 235

cag tct tca gtc aaa gca gca gac aaa agt tcg cct gct ccc ttt ttg   38203
Gln Ser Ser Val Lys Ala Ala Asp Lys Ser Ser Pro Ala Pro Phe Leu
                240                 245                 250

gtg atg ccg cct cta ttt cgt tta gaa gca aac cag cag agt caa ctg   38251
Val Met Pro Pro Leu Phe Arg Leu Glu Ala Asn Gln Gln Ser Gln Leu
            255                 260                 265

cgt att gtc cgt act ggt ggt gac atg cca acg gat cgt gag act tta   38299
Arg Ile Val Arg Thr Gly Gly Asp Met Pro Thr Asp Arg Glu Thr Leu
        270                 275                 280

cag tgg gtc tgt ata aag gcg gta cca ccc gaa aat gaa ccg tcg gat   38347
Gln Trp Val Cys Ile Lys Ala Val Pro Pro Glu Asn Glu Pro Ser Asp
    285                 290                 295

aca cag gct aag ggc gcg acc ctt gac ctc aat ttg tcc atc aac gcc   38395
Thr Gln Ala Lys Gly Ala Thr Leu Asp Leu Asn Leu Ser Ile Asn Ala
300                 305                 310                 315

tgt gat aag ctg att ttc cgc ccg gat gcc gtg aag ggg acg ccg gaa   38443
Cys Asp Lys Leu Ile Phe Arg Pro Asp Ala Val Lys Gly Thr Pro Glu
                320                 325                 330

gat gtt gca gga aat tta aga tgg gtg gag acg ggc aac aaa ctt aag   38491
Asp Val Ala Gly Asn Leu Arg Trp Val Glu Thr Gly Asn Lys Leu Lys
            335                 340                 345

gtg gag aac ccc acc ccg ttt tac atg aat tta gcc tct gtc aca gta   38539
Val Glu Asn Pro Thr Pro Phe Tyr Met Asn Leu Ala Ser Val Thr Val
        350                 355                 360

ggg gga aag ccc att aca ggg ctt gag tat gtc ccc ccc ttt gct gac   38587
Gly Gly Lys Pro Ile Thr Gly Leu Glu Tyr Val Pro Pro Phe Ala Asp
    365                 370                 375

aaa aca cta aat atg cca ggt agt gcc cat ggt gat atc gag tgg aga   38635
Lys Thr Leu Asn Met Pro Gly Ser Ala His Gly Asp Ile Glu Trp Arg
380                 385                 390                 395

gtt att aca gac ttt ggt ggt gaa agt cat ccg ttc cac tac gtt ctt   38683
Val Ile Thr Asp Phe Gly Gly Glu Ser His Pro Phe His Tyr Val Leu
                400                 405                 410

aaa taa atccaggggc ttagcggcag aaa atg aag ttc aaa caa cct gcc ttg 38736
Lys                             Met Lys Phe Lys Gln Pro Ala Leu
                                    415                 420

cta ctg ttc atc gcg gga gtg gtt cat tgc gca aat gcg cac act tac   38784
```

```
Leu Leu Phe Ile Ala Gly Val Val His Cys Ala Asn Ala His Thr Tyr
            425                 430                 435

aca ttc gat gca tca atg ttg ggc gat gca gcg aaa ggg gtt gat atg   38832
Thr Phe Asp Ala Ser Met Leu Gly Asp Ala Ala Lys Gly Val Asp Met
        440                 445                 450

tcg ctc ttt aac cag ggg tta caa cag cca ggg act tat cgc gtg gac   38880
Ser Leu Phe Asn Gln Gly Leu Gln Gln Pro Gly Thr Tyr Arg Val Asp
    455                 460                 465

gtg atg gtg aat ggt aaa cgt gtc gac acc cgt gat gtg gtg ttc aaa   38928
Val Met Val Asn Gly Lys Arg Val Asp Thr Arg Asp Val Val Phe Lys
470                 475                 480                 485

ttg gaa aag gat ggg caa gga acg cct gtt ctg gct cct tgt ttg acg   38976
Leu Glu Lys Asp Gly Gln Gly Thr Pro Val Leu Ala Pro Cys Leu Thr
                490                 495                 500

gtc agt cag ctt tca cgc tac ggc gta aaa acg gaa gat tac cct cag   39024
Val Ser Gln Leu Ser Arg Tyr Gly Val Lys Thr Glu Asp Tyr Pro Gln
            505                 510                 515

ttg tgg aaa gca gca aag ccc cca gat gag tgt gcg gat ctg acc gcc   39072
Leu Trp Lys Ala Ala Lys Pro Pro Asp Glu Cys Ala Asp Leu Thr Ala
        520                 525                 530

att cca cag gct aaa gcg gta ctg gat atc aat aat cag caa ctg caa   39120
Ile Pro Gln Ala Lys Ala Val Leu Asp Ile Asn Asn Gln Gln Leu Gln
    535                 540                 545

ctg agt att ccg cag ttg gcg ttg cgt ccg gaa ttt aag ggg atc gct   39168
Leu Ser Ile Pro Gln Leu Ala Leu Arg Pro Glu Phe Lys Gly Ile Ala
550                 555                 560                 565

cca gaa gat ctt tgg gat gat ggt att ccg gcg ttt ctg atg aac tac   39216
Pro Glu Asp Leu Trp Asp Asp Gly Ile Pro Ala Phe Leu Met Asn Tyr
                570                 575                 580

agt gcg agg aca acg cag acg gat tac aaa atg gat atg gtg ggg cgt   39264
Ser Ala Arg Thr Thr Gln Thr Asp Tyr Lys Met Asp Met Val Gly Arg
            585                 590                 595

gac aac tct tcc tgg gta caa ctg caa ccg gga atc aat ata ggt gcg   39312
Asp Asn Ser Ser Trp Val Gln Leu Gln Pro Gly Ile Asn Ile Gly Ala
        600                 605                 610

tgg cgt gtc cgc aat gcg acc agc tgg cag cgg agt agt caa ctg tcg   39360
Trp Arg Val Arg Asn Ala Thr Ser Trp Gln Arg Ser Ser Gln Leu Ser
    615                 620                 625

ggg aag tgg cag gca gca tat acc tat gct gag cgt gga ctg tac tca   39408
Gly Lys Trp Gln Ala Ala Tyr Thr Tyr Ala Glu Arg Gly Leu Tyr Ser
630                 635                 640                 645

cta aaa agt cgt ctg act ctg ggg caa aag act tcg cag ggg gag ata   39456
Leu Lys Ser Arg Leu Thr Leu Gly Gln Lys Thr Ser Gln Gly Glu Ile
                650                 655                 660

ttt gat agt gtg cca ttt acc ggt gtg atg ttg gca tcg gat gac aac   39504
Phe Asp Ser Val Pro Phe Thr Gly Val Met Leu Ala Ser Asp Asp Asn
```

```
                665                 670                 675

atg gtg ccc tac agt gag cgg cag ttt gct ccg gta gtg cgt ggg att   39552
Met Val Pro Tyr Ser Glu Arg Gln Phe Ala Pro Val Val Arg Gly Ile
        680                 685                 690

gcc cgc acg cag gct cgg gtg gag gtc aaa cag aat ggt tac acc att   39600
Ala Arg Thr Gln Ala Arg Val Glu Val Lys Gln Asn Gly Tyr Thr Ile
    695                 700                 705

tac aac acc act gtg gcg ccc gga ccg ttt gca ctg cgg gat ctg tcg   39648
Tyr Asn Thr Thr Val Ala Pro Gly Pro Phe Ala Leu Arg Asp Leu Ser
710                 715                 720                 725

gta aca gac agt agt ggt gat ctg cat gtc acc gtg tgg gag gcc gat   39696
Val Thr Asp Ser Ser Gly Asp Leu His Val Thr Val Trp Glu Ala Asp
                730                 735                 740

ggc agt aca caa atg ttt gtg gtg ccg tat cag acc ccg gcg ata gca   39744
Gly Ser Thr Gln Met Phe Val Val Pro Tyr Gln Thr Pro Ala Ile Ala
            745                 750                 755

ctg cac cag gga tat ttg aag tac agc ctg ttg gcg ggc cga tac cga   39792
Leu His Gln Gly Tyr Leu Lys Tyr Ser Leu Leu Ala Gly Arg Tyr Arg
        760                 765                 770

tcg tca gac tct gca acg gat aag cgg cag atc gcg cag gct acg ttg   39840
Ser Ser Asp Ser Ala Thr Asp Lys Arg Gln Ile Ala Gln Ala Thr Leu
    775                 780                 785

atg tat ggt ctg ccg tgg aat ctc act gca tac ggc ggt ata cag agt   39888
Met Tyr Gly Leu Pro Trp Asn Leu Thr Ala Tyr Gly Gly Ile Gln Ser
790                 795                 800                 805

gca acg cat aat caa gct gca ttg ctt ggt ttg ggg gga tct ctc ggg   39936
Ala Thr His Asn Gln Ala Ala Leu Leu Gly Leu Gly Gly Ser Leu Gly
                810                 815                 820

cgg tgg ggg agt tta tct gtc gat gga agc gac aca cac agt cag cgt   39984
Arg Trp Gly Ser Leu Ser Val Asp Gly Ser Asp Thr His Ser Gln Arg
            825                 830                 835

cag ggg gag gcg gta cag caa gga gcc tcc tgg cga ctg cgt tac agc   40032
Gln Gly Glu Ala Val Gln Gln Gly Ala Ser Trp Arg Leu Arg Tyr Ser
        840                 845                 850

aac cag ctg act gcg acg ggg aca aat ttt ttt ctg acg aga tgg cag   40080
Asn Gln Leu Thr Ala Thr Gly Thr Asn Phe Phe Leu Thr Arg Trp Gln
    855                 860                 865

tat gcc tcg cag ggc tat aac acc cta tcc gat gtg ctc gac agt tat   40128
Tyr Ala Ser Gln Gly Tyr Asn Thr Leu Ser Asp Val Leu Asp Ser Tyr
870                 875                 880                 885

cga cat aat ggc aac cgt cta tgg tcg tgg cgg gaa aat ttg cag ccg   40176
Arg His Asn Gly Asn Arg Leu Trp Ser Trp Arg Glu Asn Leu Gln Pro
                890                 895                 900

agc tcg cgt act acc ctg atg ttg agt cag tca tgg ggg agg cat ttg   40224
Ser Ser Arg Thr Thr Leu Met Leu Ser Gln Ser Trp Gly Arg His Leu
            905                 910                 915
```

```
ggc aat ctg agt tta acc ggt tcc cgt acc gac tgg cgt aat cgc ccc   40272
Gly Asn Leu Ser Leu Thr Gly Ser Arg Thr Asp Trp Arg Asn Arg Pro
        920                 925                 930

ggt cat gat gac agc tac gga ctg agt tgg gga acc tct atc gga ggg   40320
Gly His Asp Asp Ser Tyr Gly Leu Ser Trp Gly Thr Ser Ile Gly Gly
    935                 940                 945

ggc tcg ctg tca ttg aac tgg aat caa aac aga acg ctg tgg cgc aat   40368
Gly Ser Leu Ser Leu Asn Trp Asn Gln Asn Arg Thr Leu Trp Arg Asn
950                 955                 960                 965

ggc gcg cac cgt aaa gag aac ata acc agc ctg tgg ttc agt atg cca   40416
Gly Ala His Arg Lys Glu Asn Ile Thr Ser Leu Trp Phe Ser Met Pro
                970                 975                 980

tta agc cgc tgg acg ggg aat aat gta agt gct agt tgg cag atg act   40464
Leu Ser Arg Trp Thr Gly Asn Asn Val Ser Ala Ser Trp Gln Met Thr
            985                 990                 995

tca cca tca cac ggt ggt cag acg caa caa gtg ggg gtc aac gga gag   40512
Ser Pro Ser His Gly Gly Gln Thr Gln Gln Val Gly Val Asn Gly Glu
       1000                1005                1010

gca ttc agt cag caa ctg gat tgg gag gtg cgt cag agt tac cgt gcc   40560
Ala Phe Ser Gln Gln Leu Asp Trp Glu Val Arg Gln Ser Tyr Arg Ala
   1015                1020                1025

gat gcc ccg cca ggt ggt ggt aat aac agc gca ttg cac ttg gca tgg   40608
Asp Ala Pro Pro Gly Gly Gly Asn Asn Ser Ala Leu His Leu Ala Trp
1030                1035                1040                1045

aat ggg gat tac ggc ctg tta ggt ggt gac tat agc tac agc cgg gcg   40656
Asn Gly Asp Tyr Gly Leu Leu Gly Gly Asp Tyr Ser Tyr Ser Arg Ala
               1050                1055                1060

atg cgc cag atg gga gtc aat atc gcg gga ggt ata gtt atc cac cat   40704
Met Arg Gln Met Gly Val Asn Ile Ala Gly Gly Ile Val Ile His His
           1065                1070                1075

cat ggt gtg acg ctg ggg caa cct ttg caa ggc tca gtg gcg ctg gtt   40752
His Gly Val Thr Leu Gly Gln Pro Leu Gln Gly Ser Val Ala Leu Val
       1080                1085                1090

gaa gcg cca ggg gcc tcg ggg gtg cca gtt ggc ggc tgg cct ggc gtt   40800
Glu Ala Pro Gly Ala Ser Gly Val Pro Val Gly Gly Trp Pro Gly Val
   1095                1100                1105

aag acg gat ttt cgt ggc gac acc aca gtg ggc aac ctg aac gtc tat   40848
Lys Thr Asp Phe Arg Gly Asp Thr Thr Val Gly Asn Leu Asn Val Tyr
1110                1115                1120                1125

cag gag aat aca gtc agc ctc gat ccg tcg cga cta ccg gat gac gca   40896
Gln Glu Asn Thr Val Ser Leu Asp Pro Ser Arg Leu Pro Asp Asp Ala
               1130                1135                1140

gag gtc aca caa acc gat gtg cgc gtg gtg cca acc gaa ggg gcg gtg   40944
Glu Val Thr Gln Thr Asp Val Arg Val Val Pro Thr Glu Gly Ala Val
           1145                1150                1155
```

```
gtg gaa gcg aag ttt cac act cgc atc ggg gcc agg gca ctg atg acg   40992
Val Glu Ala Lys Phe His Thr Arg Ile Gly Ala Arg Ala Leu Met Thr
       1160                1165                1170

ctg aaa cgg gaa gat ggt agc gcc att cct ttc ggg gcg cag gtt aca   41040
Leu Lys Arg Glu Asp Gly Ser Ala Ile Pro Phe Gly Ala Gln Val Thr
   1175                1180                1185

gtc aat ggg cag gat ggc agt gct gct ctg gtg gat act gat agc cag   41088
Val Asn Gly Gln Asp Gly Ser Ala Ala Leu Val Asp Thr Asp Ser Gln
1190                1195                1200                1205

gtt tat ctc act ggt ttg gcg gat aag ggc gaa ctg acg gtg aaa tgg   41136
Val Tyr Leu Thr Gly Leu Ala Asp Lys Gly Glu Leu Thr Val Lys Trp
               1210                1215                1220

gga gca cag caa tgt cgg gtt aac tac cgc cta cct gcc cac aag gga   41184
Gly Ala Gln Gln Cys Arg Val Asn Tyr Arg Leu Pro Ala His Lys Gly
           1225                1230                1235

atc gcg ggc ttg tat caa atg agc ggt ctc tgc aga tag ccgattctga   41233
Ile Ala Gly Leu Tyr Gln Met Ser Gly Leu Cys Arg
       1240                1245                1250

aggagagaat a atg tgg atg aaa ata cag cga gtg aaa acg gtt atc tat  41283
             Met Trp Met Lys Ile Gln Arg Val Lys Thr Val Ile Tyr
                         1255                1260

agc gta agc tta ctg gtc gct gcc agt agc ttg gtg ccg ata gcg aac   41331
Ser Val Ser Leu Leu Val Ala Ala Ser Ser Leu Val Pro Ile Ala Asn
   1265                1270                1275

gcc gca gaa aaa ctt cag aca acg cta cgt gta ggt act tac ttt cgg   41379
Ala Ala Glu Lys Leu Gln Thr Thr Leu Arg Val Gly Thr Tyr Phe Arg
1280                1285                1290                1295

gct ggg cac gtg cca gat ggg atg gtg ctt gcg caa ggc tgg gtg act   41427
Ala Gly His Val Pro Asp Gly Met Val Leu Ala Gln Gly Trp Val Thr
               1300                1305                1310

tat cac ggc agt cac agc ggg ttt cgg gta tgg agc gat gag caa aag   41475
Tyr His Gly Ser His Ser Gly Phe Arg Val Trp Ser Asp Glu Gln Lys
           1315                1320                1325

gcg ggt aac acg cct acc gta ttg ctg ctg agc ggg caa cag gat cct   41523
Ala Gly Asn Thr Pro Thr Val Leu Leu Leu Ser Gly Gln Gln Asp Pro
       1330                1335                1340

cgc cat cac att cag gtt cgc ctg gag ggc gag ggg tgg caa cca gat   41571
Arg His His Ile Gln Val Arg Leu Glu Gly Glu Gly Trp Gln Pro Asp
   1345                1350                1355

acg gtg agt ggt cgt ggc gcc att tta aga acc gct gca gat aac gcc   41619
Thr Val Ser Gly Arg Gly Ala Ile Leu Arg Thr Ala Ala Asp Asn Ala
1360                1365                1370                1375

agt ttc agt gtg gtc gtt gat ggc aat cag gaa gtg cct gcg gac acc   41667
Ser Phe Ser Val Val Val Asp Gly Asn Gln Glu Val Pro Ala Asp Thr
               1380                1385                1390

tgg acg ctg gat ttt aag gcc tgt gca ttg gcg cag gag gat acg tag   41715
```

Trp Thr Leu Asp Phe Lys Ala Cys Ala Leu Ala Gln Glu Asp Thr
1395 1400 1405

ccgtctgttc cactcatact tcctgtatca gtgaataagg cgaatactgt atatcaggtt 41775

cgaggatgta ttgaaaattt actcaaggca aatatggtta gcaatcttca ttgcagaaaa 41835

atcaattgct gattgatttt aaaatctgaa attcatcttt tttgtaggga gaggaatatt 41895

atgtttcgta aaataattta acggacactg atgcctttat tatttctgtt ttgtggattt 41955

gtttgattca ctcttttgtg agcggggctt tataagcccc tggtataagg tttgtttatc 42015

tgtcaggtat tcatatgtga tatttaagat tttatgttta gggggctttt tcttgtcccc 42075

tcaggcgtaa aaataattta ttttttacat aaaggaataa agcatatgtc ttatgcacga 42135

catttaccgg tattaatgta tcaccatgtc agtgataaac ccggacagat aaccttatct 42195

ccccgtacgt tccgggcgca gatgaaatgg ctggccgaat ctggctggaa aaccgttact 42255

gcctgcagag gtggaagctt tttatcatgg tgcaagattg cctcgtaaaa gcgtcatgct 42315

gacctttgat ggcggctggc tggataactg gttgcaggtt tttccggtgc tgcaggagtt 42375

taatctgcat gcgcatctct ttcttgtgac cagtttgatc agtgacggac cggtccgtat 42435

tcctgcaggc gaaccggtgt actctcatga tgagtgtcaa atgctggtta aacaaggccg 42495

ggctgatgag gtcatgctgc gctggtcaga ggtccgggag atgcacctca gtggccttgt 42555

tgagtttcac tcgcacacgc acacccaccg acgctgggac cagaagcctg tgtcccgtaa 42615

tccgtcggat ttgcttcgtg tcgatattct tcttagtcgt aagcggatga gggagatgct 42675

gggttattgc agtcagcatc tgtgctggcc tgagggctgg tattgttctg actatattca 42735

tgtggctgaa gagttggggt tcacatacct gtataccaca gaaaggcgta tgaacaatcc 42795

agtcatcggt tcacagcgta ttggtcgtat caacgcaaag gagcgaaaga atgtgggctg 42855

gctgaaacgt cgtctgtttt atcacaccac gcccggattt tcttcgctgc tggcccggca 42915

taaggggggca cgtcggatag ctgactgagc cagagaccag gatgaaagtg ctgcataccg 42975

aatcctcccc catcatcggc gggcaggggt tgtaggctat atcccaaatg atggtgctga 43035

ttttgagcag aaataagata acaccggagg ccgtgataga cccggttgcg cagaatctgg 43095

aggcagcagg tcactggcgc agggcttcta cccgctggct tttggttatg ggagattttg 43155

aatgtacaga ggcccagcgg gagtggttgt tgttgcgccg gaattattgc cttgcgcaga 43215

tatcttcccc tggcgccagt aaagctggat atcagcgacg tgacgaaatg agaggagcct 43275

gcacataggt ttgtgttttg agtaagtgtc ttatttaaac cgtctgttct gtttcctccg 43335

ctttcacaaa taatgtcgag ccgggtgggg gactcaagta agaataatct ggcgatgttt 43395

tgcttgtttc cacgggatac tttgttaggt gaacgataca attaatgcgc tccagactgc 43455

gtcacagaaa aggagaatgc ctttgcagaa cgggcgcatg acaatgaacc actttattat 43515

tggtatctgg ttcggcgaca accttgacga tcacgtctga gatagggtat ttacactgag 43575

tggtaaacag gttcaattag taaccggaga tggatgcaaa atcatgatcg attcagatgc 43635

tattttgcag ccaatagatt ttttattaag atgatatcaa ggacattgag gcacacaatg 43695

acgtatcagt aagtcgttga tagctcattt gatataagaa tttcttttat caacggaaga 43755

taatgatgga actgatcaat aatcgtggta tgcgagactg gatgatattt attaaagtgg 43815

cggaagtagg gaatctttcc cgggctgcgc gggaattaga tattagcatt tctgctgtca 43875

gtaaatcgct tagtcgcctt gagaattcta ttgaggttac tttacttcgg cgcgattcac 43935

atcacttaga actgactgga gctggtcaga cagcctatgc aagcatgaaa aggataacat 43995

cttcctttca gtccttgctg gatgaattgc gaaatccgga taaaattatc agagggagta 44055

taaaattttc ggctccggct attgtctgtg agtttcttgc caataagtgg atatgggaat 44115

ttacagctag ctatccggat acaaaaatct acctggattc acgagagcgt agcgattttt 44175

ttagtaaatc cctggagttt gatgagctgg tttttaaaag tggcataatc gaaagtgagg 44235

atctcgtgta tcgaaagata agccctttaa agttggttct ttgtgcgagt ccgaaatata 44295

tcagaaaata tggcaggatc tcacaccctg gcgatttgga aaatcacatt attgtgggtc 44355

ttcacaacca tggtctttcc ggacctctta ctcttttccg tcaggatgaa tcatacacta 44415

ttagtggcgc tgttaatgtt catttatctt ccaataatct tttgagtgtt cttaatttgg 44475

ttttagaagg aaagggtatc aacctcatga ctccggcctg gcttgccacc aaatacttaa 44535

aaaataatga acttgaaatt atacttcctg aatggagggt tccagatctc cccatttatc 44595

ttgtatggcg tcatcgtcag tattattctc ctttatttca acgctttctg tcttttattg 44655

aagataaatg gaataatcgc ccacaaattg attttctgaa tgatgattaa cccgtttgga 44715

atggttttga tacgttcctg acttaaaccc acatgatgac tgaattgagg catcgagata 44775

tgcgactggt cagccagtcg tcttttgacg atgcccaata caaaacatga tgccgactga 44835

cggaaatgat aatacgcgga aacaggacgg ggctgttttt gggcagccgg aagttaagcc 44895

cataccagaa acgttgcagt gtactgaaaa atggcgccag gttgcacctg ttcaaagatt 44955

ttctgaaggc gcaggagtat tcattactga tatcctccat tgcgccttcg ggaacccaca 45015

ggaccagcta ttttaccgat agtgtttaaa aggcgtaagt aatgccgagc atgaagtcat 45075

tggaggcagc ctttgtgtct gcatcataag cggtatgttc atcaccagca tagtgatttt 45135

ttgaaatgct tactttgcca gcattaatgt atttataact ggcgtcaatc ataatattat 45195

ctgttacagc atattttgca ccgatacctg cgccccaggc aaagttattt tttgaagcag 45255

acagagtttc attaatacca aaaccaacag gaatggtgtt attacttagc ttcacatgag 45315

cgaggccaac gcctgcgctg atatagggag taaatgccgt actattgtga aaatcataat 45375

agccattaac catgtaagtg gtcattcgga cctgattttt tacatttatg tgtactggat 45435

caccaaatgc aataatatcc tgcccgcctt tagcatccgt ctcacctctg aaagtggtat 45495

ccagttctaa acgtactgga agctggaatg gatcataaaa gtcataaccg atagcaaccc 45555

cgccgccaaa aacgcctttg gtacggtcag gtaacgttgc atgaccatta actatctcat 45615

cctggctgaa ggttgagttg attccataga cattgactac ggatgtcccc gctttcccgg 45675

tgatatagat cccttctttt gctgatgcag tagcggacca ggctaccaca aggggaatga 45735

tgcagactgc gaaaaagttt ttcatttcag aacctgcctt aatattgggc taaaagacaa 45795

gtttcacggt atagggtgtg atataacgat tacataaacg aagcccaaaa aacggtctat 45855

tgtaacgctg ggttttctgt aagcgggtaa aaaatgagat gaagatttta aataacaata 45915

cgataatcgt cggtatggaa atccatctcc tcgccaaatt gccccacgta cggtttcact 45975

tctacgttat gtaacgggta gtgtgagatg gagcgatgct gtaagaaaaa gatgaagatg 46035

aatttgtacc cgacctggat aaagcccgtt atcccggaat aacgggcaaa aatatttact 46095

caagtgcctg ggcgagatct tgttgtacct gttgacgctg ttctggtgtt aagactttgc 46155

ttaaatcaaa ataatattta acccgataat agcgagcctg ttgttctatg ttactgaagg 46215

ctgcaagctg ctgttttacg gcggcgtcat cccatttacc ggatttaatc acctctatca 46275

gcgcaccgtc tttaattccc ttcatagaaa tctgactgac gtcggtttcc agttgttggt 46335

gaagtttttt gatccgggta atctgatcgt ttgtcagctt cagatgctgg acaataggat 46395

cctgggcggg caggggagga ttggggacag cggtggcgaa agcgccaaaa gaaacgcccg 46455

ccagagtcgc tgccagtaaa gttgtgcgta caaagttttt catgaagata tcctgataag 46515

ggagtgatta accgttttta ttacccacga atggcgagca attatcttag agcctatccc 46575

agtagggcta ttttacttgc cattttggac ctgggcagtg ctcgccaaaa cgcgttagcg 46635

ttttgaacgc cgctagcggc ggcccgaagg gcgagcgtag cgagtcaaac ctcacgtact 46695

acgtgtacgc tccggttttt gcgcgctgtc cgtgtccaaa ctggctgcgc caataacgcc 46755

tggtgggata ggctcttagt cagaatacgt tgcctgccac attacgccac gcgaatttgt 46815

tttacggaga gttacggagt gaaacaatcc cgccgcggtg agcggcaggt tgctt 46870

<210> 2
<211> 166
<212> PRT
<213> Salmonella typhimurium

<400> 2
Met Lys Ser Ile Lys Lys Leu Ile Ile Ala Ser Ala Leu Ser Met Met

```
  1               5                  10                  15

Ala Ala Ser Cys Tyr Ala Gly Ser Phe Leu Pro Asn Ser Glu Gln Gln
            20                  25                  30

Lys Ser Val Asp Ile Val Phe Ser Ser Pro Gln Asp Leu Thr Val Ser
        35                  40                  45

Leu Ile Pro Val Ser Gly Leu Lys Ala Gly Lys Asn Ala Pro Ser Ala
    50                  55                  60

Lys Ile Ala Lys Leu Val Val Asn Ser Thr Thr Leu Lys Glu Phe Gly
65                  70                  75                  80

Val Arg Gly Ile Ser Asn Asn Val Val Asp Ser Thr Gly Thr Ala Trp
                85                  90                  95

Arg Val Ala Gly Lys Asn Thr Gly Lys Glu Ile Gly Val Gly Leu Ser
            100                 105                 110

Ser Asp Ser Leu Arg Arg Ser Asp Ser Thr Glu Lys Trp Asn Gly Val
        115                 120                 125

Asn Trp Met Thr Phe Asn Ser Asn Asp Thr Leu Asp Ile Val Leu Thr
    130                 135                 140

Gly Pro Ala Gln Asn Val Thr Ala Asp Thr Tyr Pro Ile Thr Leu Asp
145                 150                 155                 160

Val Val Gly Tyr Gln Pro
            165


<210> 3
<211> 245
<212> PRT
<213> Salmonella typhimurium

<400> 3
Met Lys Ile Val Asn Phe Ala Val Met Ala Val Ala Leu Phe Ala Thr
  1               5                  10                  15

Asn Ser Met Val Ser Val Tyr Ala Val Asn Gln Gln Leu Asn Ser Ala
            20                  25                  30

Thr Lys Leu Phe Ser Val Lys Leu Gly Ala Thr Arg Val Ile Tyr His
        35                  40                  45

Ala Gly Thr Ala Gly Ala Thr Leu Ser Val Ser Asn Pro Gln Asn Tyr
    50                  55                  60

Pro Ile Leu Val Gln Ser Ser Val Lys Ala Ala Asp Lys Ser Ser Pro
65                  70                  75                  80

Ala Pro Phe Leu Val Met Pro Pro Leu Phe Arg Leu Glu Ala Asn Gln
                85                  90                  95

Gln Ser Gln Leu Arg Ile Val Arg Thr Gly Gly Asp Met Pro Thr Asp
            100                 105                 110

Arg Glu Thr Leu Gln Trp Val Cys Ile Lys Ala Val Pro Pro Glu Asn
```

```
        115                 120                 125

Glu Pro Ser Asp Thr Gln Ala Lys Gly Ala Thr Leu Asp Leu Asn Leu
    130                 135                 140

Ser Ile Asn Ala Cys Asp Lys Leu Ile Phe Arg Pro Asp Ala Val Lys
145                 150                 155                 160

Gly Thr Pro Glu Asp Val Ala Gly Asn Leu Arg Trp Val Glu Thr Gly
                165                 170                 175

Asn Lys Leu Lys Val Glu Asn Pro Thr Pro Phe Tyr Met Asn Leu Ala
            180                 185                 190

Ser Val Thr Val Gly Gly Lys Pro Ile Thr Gly Leu Glu Tyr Val Pro
        195                 200                 205

Pro Phe Ala Asp Lys Thr Leu Asn Met Pro Gly Ser Ala His Gly Asp
    210                 215                 220

Ile Glu Trp Arg Val Ile Thr Asp Phe Gly Gly Glu Ser His Pro Phe
225                 230                 235                 240

His Tyr Val Leu Lys
                245
```

<210> 4
<211> 836
<212> PRT
<213> Salmonella typhimurium

```
<400> 4
Met Lys Phe Lys Gln Pro Ala Leu Leu Leu Phe Ile Ala Gly Val Val
  1               5                  10                  15

His Cys Ala Asn Ala His Thr Tyr Thr Phe Asp Ala Ser Met Leu Gly
            20                  25                  30

Asp Ala Ala Lys Gly Val Asp Met Ser Leu Phe Asn Gln Gly Leu Gln
        35                  40                  45

Gln Pro Gly Thr Tyr Arg Val Asp Val Met Val Asn Gly Lys Arg Val
    50                  55                  60

Asp Thr Arg Asp Val Val Phe Lys Leu Glu Lys Asp Gly Gln Gly Thr
65                  70                  75                  80

Pro Val Leu Ala Pro Cys Leu Thr Val Ser Gln Leu Ser Arg Tyr Gly
                85                  90                  95

Val Lys Thr Glu Asp Tyr Pro Gln Leu Trp Lys Ala Ala Lys Pro Pro
            100                 105                 110

Asp Glu Cys Ala Asp Leu Thr Ala Ile Pro Gln Ala Lys Ala Val Leu
        115                 120                 125

Asp Ile Asn Asn Gln Gln Leu Gln Leu Ser Ile Pro Gln Leu Ala Leu
    130                 135                 140

Arg Pro Glu Phe Lys Gly Ile Ala Pro Glu Asp Leu Trp Asp Asp Gly
```

```
145                 150                 155                 160

Ile Pro Ala Phe Leu Met Asn Tyr Ser Ala Arg Thr Thr Gln Thr Asp
                165                 170                 175

Tyr Lys Met Asp Met Val Gly Arg Asp Asn Ser Ser Trp Val Gln Leu
            180                 185                 190

Gln Pro Gly Ile Asn Ile Gly Ala Trp Arg Val Arg Asn Ala Thr Ser
        195                 200                 205

Trp Gln Arg Ser Ser Gln Leu Ser Gly Lys Trp Gln Ala Ala Tyr Thr
    210                 215                 220

Tyr Ala Glu Arg Gly Leu Tyr Ser Leu Lys Ser Arg Leu Thr Leu Gly
225                 230                 235                 240

Gln Lys Thr Ser Gln Gly Glu Ile Phe Asp Ser Val Pro Phe Thr Gly
                245                 250                 255

Val Met Leu Ala Ser Asp Asp Asn Met Val Pro Tyr Ser Glu Arg Gln
            260                 265                 270

Phe Ala Pro Val Val Arg Gly Ile Ala Arg Thr Gln Ala Arg Val Glu
        275                 280                 285

Val Lys Gln Asn Gly Tyr Thr Ile Tyr Asn Thr Thr Val Ala Pro Gly
    290                 295                 300

Pro Phe Ala Leu Arg Asp Leu Ser Val Thr Asp Ser Ser Gly Asp Leu
305                 310                 315                 320

His Val Thr Val Trp Glu Ala Asp Gly Ser Thr Gln Met Phe Val Val
                325                 330                 335

Pro Tyr Gln Thr Pro Ala Ile Ala Leu His Gln Gly Tyr Leu Lys Tyr
            340                 345                 350

Ser Leu Leu Ala Gly Arg Tyr Arg Ser Ser Asp Ser Ala Thr Asp Lys
        355                 360                 365

Arg Gln Ile Ala Gln Ala Thr Leu Met Tyr Gly Leu Pro Trp Asn Leu
    370                 375                 380

Thr Ala Tyr Gly Gly Ile Gln Ser Ala Thr His Asn Gln Ala Ala Leu
385                 390                 395                 400

Leu Gly Leu Gly Gly Ser Leu Gly Arg Trp Gly Ser Leu Ser Val Asp
                405                 410                 415

Gly Ser Asp Thr His Ser Gln Arg Gln Gly Glu Ala Val Gln Gln Gly
            420                 425                 430

Ala Ser Trp Arg Leu Arg Tyr Ser Asn Gln Leu Thr Ala Thr Gly Thr
        435                 440                 445

Asn Phe Phe Leu Thr Arg Trp Gln Tyr Ala Ser Gln Gly Tyr Asn Thr
    450                 455                 460

Leu Ser Asp Val Leu Asp Ser Tyr Arg His Asn Gly Asn Arg Leu Trp
465                 470                 475                 480
```

```
Ser Trp Arg Glu Asn Leu Gln Pro Ser Ser Arg Thr Thr Leu Met Leu
                485                 490                 495

Ser Gln Ser Trp Gly Arg His Leu Gly Asn Leu Ser Leu Thr Gly Ser
            500                 505                 510

Arg Thr Asp Trp Arg Asn Arg Pro Gly His Asp Asp Ser Tyr Gly Leu
        515                 520                 525

Ser Trp Gly Thr Ser Ile Gly Gly Gly Ser Leu Ser Leu Asn Trp Asn
    530                 535                 540

Gln Asn Arg Thr Leu Trp Arg Asn Gly Ala His Arg Lys Glu Asn Ile
545                 550                 555                 560

Thr Ser Leu Trp Phe Ser Met Pro Leu Ser Arg Trp Thr Gly Asn Asn
                565                 570                 575

Val Ser Ala Ser Trp Gln Met Thr Ser Pro Ser His Gly Gly Gln Thr
            580                 585                 590

Gln Gln Val Gly Val Asn Gly Glu Ala Phe Ser Gln Gln Leu Asp Trp
        595                 600                 605

Glu Val Arg Gln Ser Tyr Arg Ala Asp Ala Pro Pro Gly Gly Gly Asn
    610                 615                 620

Asn Ser Ala Leu His Leu Ala Trp Asn Gly Asp Tyr Gly Leu Leu Gly
625                 630                 635                 640

Gly Asp Tyr Ser Tyr Ser Arg Ala Met Arg Gln Met Gly Val Asn Ile
                645                 650                 655

Ala Gly Gly Ile Val Ile His His His Gly Val Thr Leu Gly Gln Pro
            660                 665                 670

Leu Gln Gly Ser Val Ala Leu Val Glu Ala Pro Gly Ala Ser Gly Val
        675                 680                 685

Pro Val Gly Gly Trp Pro Gly Val Lys Thr Asp Phe Arg Gly Asp Thr
    690                 695                 700

Thr Val Gly Asn Leu Asn Val Tyr Gln Glu Asn Thr Val Ser Leu Asp
705                 710                 715                 720

Pro Ser Arg Leu Pro Asp Asp Ala Glu Val Thr Gln Thr Asp Val Arg
                725                 730                 735

Val Val Pro Thr Glu Gly Ala Val Val Glu Ala Lys Phe His Thr Arg
            740                 745                 750

Ile Gly Ala Arg Ala Leu Met Thr Leu Lys Arg Glu Asp Gly Ser Ala
        755                 760                 765

Ile Pro Phe Gly Ala Gln Val Thr Val Asn Gly Gln Asp Gly Ser Ala
    770                 775                 780

Ala Leu Val Asp Thr Asp Ser Gln Val Tyr Leu Thr Gly Leu Ala Asp
785                 790                 795                 800
```

```
Lys Gly Glu Leu Thr Val Lys Trp Gly Ala Gln Gln Cys Arg Val Asn
                805                 810                 815

Tyr Arg Leu Pro Ala His Lys Gly Ile Ala Gly Leu Tyr Gln Met Ser
            820                 825                 830

Gly Leu Cys Arg
        835
```

<210> 5
<211> 156
<212> PRT
<213> Salmonella typhimurium

<400> 5

```
Met Trp Met Lys Ile Gln Arg Val Lys Thr Val Ile Tyr Ser Val Ser
  1               5                  10                  15

Leu Leu Val Ala Ala Ser Ser Leu Val Pro Ile Ala Asn Ala Ala Glu
             20                  25                  30

Lys Leu Gln Thr Thr Leu Arg Val Gly Thr Tyr Phe Arg Ala Gly His
         35                  40                  45

Val Pro Asp Gly Met Val Leu Ala Gln Gly Trp Val Thr Tyr His Gly
     50                  55                  60

Ser His Ser Gly Phe Arg Val Trp Ser Asp Glu Gln Lys Ala Gly Asn
 65                  70                  75                  80

Thr Pro Thr Val Leu Leu Leu Ser Gly Gln Gln Asp Pro Arg His His
                 85                  90                  95

Ile Gln Val Arg Leu Glu Gly Glu Gly Trp Gln Pro Asp Thr Val Ser
            100                 105                 110

Gly Arg Gly Ala Ile Leu Arg Thr Ala Ala Asp Asn Ala Ser Phe Ser
        115                 120                 125

Val Val Val Asp Gly Asn Gln Glu Val Pro Ala Asp Thr Trp Thr Leu
    130                 135                 140

Asp Phe Lys Ala Cys Ala Leu Ala Gln Glu Asp Thr
145                 150                 155
```

SEQUENCE LISTING NO. 2

<110> Folkesson, Anders

<120> The complete sequence of the tcf insert of Salmonella enterica serovar Typhi.

<130> The tcf insert in Salmonella typhi

<140>
<141>

<160> 6

<170> PatentIn Ver. 2.1

<210> 1
<211> 9253
<212> DNA
<213> Salmonella typhi

<220>
<221> CDS
<222> (1898)..(2608)
<223> tcfA putative fimbrial subunit

<220>
<221> CDS
<222> (2659)..(3234)
<223> tcfB putative fimbrial subunit

<220>
<221> CDS
<222> (3360)..(6029)
<223> tcfC putative fimbrial subunit

<220>
<221> CDS
<222> (6052)..(7131)
<223> tcfD putative fimbrial subunit

<220>
<221> CDS
<222> (7264)..(7719)
<223> tinR putative transcriptional regulator

<400> 1
tgtaagtatc ccgcataatc gagccattca catttagaga tcatccggca taatcaatct 60

gccaacgcag gagatcgctg tgcgtaaagc ccgtattact gcgcaccaga tcatcgctgt 120

gattagatca gttgaatccg gacggactgt taaagatgtc taccgggagg ccggtatttc 180

tgaagccacc agggacaact ggaagtctgg atacggcggc atggaagctt ctgatattaa 240

atcttgagga tgtcaacgcc aggatttatg gtcgtttttc gctattttta atatccgctg 300

tttgatcact tctgctgtcc gctttccgcc atttcatctt cactgattgg cgttgcgctt 360

ttggtcagcg ccccgacttt gcgtttttcc ttcagttggt aactttctcc tttgatattc 420

```
agtgtggttg agtgatgcag tagccgatcc aggatcgctg ttgccagcac gttatcaccg 480

aacatctctc cccagtctgc gaagcctttg tttgacgtca ggataatgct cgctttttca 540

taccgtcggt tcagcagtcg gaagaacaga ctggcttcct cactggtcat tggcaggtaa 600

cctatttcgt ccaggatcag cccccgcgca tagctcagtt gttgtagctg gcgctccagg 660

cggttttcca gcttcgcttt catcagcgtc taccaacctg tcctgcggca tgaacaacac 720

ccgatggccc acatctgccg ctttcacacc gggggcagcg gccaggtggg ttttctccac 780

tccgggggcc ccagcaggat cacattctcg cagcgctcca cgaacgccag accggccaac 840

tcccggacga ccttacgatc gatgcctggc tggaagctga agtcgaactg ctccagcgtt 900

ttgactcatg gcagacgagc ctgtttcagc cgggattcca ttccgcgctg atgtctgccg 960

ttctattcct gctgcagcgc catgcacagg aatgtcgact tattttgggc tgacagaggc 1020

attggtcgtt aagacttgta tattcatgtt gcttactcgc acaaaagggg gaaaggaacc 1080

gtaaacaaca ttgacaggta agggactggc atcaatttga atggtatttt ttataacatt 1140

atgttttaaa aggtgaattg tattttcatg gtgagttgtt ctgttttata ttgttgtttg 1200

tggaatattt agctaatatt taatctaaaa atagtgtatt atattgatta cactttgctg 1260

gagaggtgta aacatcaagc gtaaccatat gatgtatata agtttttgtt tgctgatatc 1320

atatttttaa cagatattga gcaagtaaat ctgatctaac tgtaacttta gttttctaaa 1380

ttaattactt tgcttacttt tttttacctg gttttgtaaa acctcatgat gatcagtcta 1440

tttgtggctt tgtttaaggt ttatttggtt ttgtagtctg aattgtatct ccctcctgga 1500

gggctttctt tagattctct gtctcctacg ttgttatgtt acgtatttgt tgctttgaag 1560

gaggggaaa tacagttcca tttatctgag taagtcaggt acacagtaac aactttctta 1620

tgaagaattt ccaaaatttt tactgcggcg ttattaattg ttcagcgatt cttacagatc 1680

tgtcgttcgc ttttggtgaa tgaaatccgt ggacttttat ttactaattt tttctttcct 1740

gaaaaaaaca gaggtattga gcgaaaaatt ttattccgta tgatgccctc cacacaaaat 1800

gtattaacac tgaatcgtaa tttgcttctt tatgctgata actttctgtc tatgctaata 1860

ctaaaattta gatgactttt atacggtaaa atctggt atg aat ttt aaa gat act  1915
                                        Met Asn Phe Lys Asp Thr
                                          1               5

ctt ccc ggg gtg ttt ctc tgt gtc gct atg ttt gca tgt ggt cat gcc  1963
Leu Pro Gly Val Phe Leu Cys Val Ala Met Phe Ala Cys Gly His Ala
            10                  15                  20

agg gcg aat atg ctc gtt tat ccc atg gcg gca gaa att aat agt agc  2011
Arg Ala Asn Met Leu Val Tyr Pro Met Ala Ala Glu Ile Asn Ser Ser
        25                  30                  35
```

```
cgc gaa gag gcc acc tcg ctg ttc gtc tat tct aaa tca gat cat gtg   2059
Arg Glu Glu Ala Thr Ser Leu Phe Val Tyr Ser Lys Ser Asp His Val
     40                  45                  50

caa tat att cga aca aga atc atg cgt att gaa cac ccc ggt atg cca   2107
Gln Tyr Ile Arg Thr Arg Ile Met Arg Ile Glu His Pro Gly Met Pro
 55                  60                  65                  70

cag gag aag gag gta cca gca ggg aat gat ata gag aca gga ctt gtt   2155
Gln Glu Lys Glu Val Pro Ala Gly Asn Asp Ile Glu Thr Gly Leu Val
                 75                  80                  85

gtc tcc ccg gag aaa ttt gct ctt tcc ccg gga aca aaa aaa aca ata   2203
Val Ser Pro Glu Lys Phe Ala Leu Ser Pro Gly Thr Lys Lys Thr Ile
             90                  95                 100

cgt gtt atc agt act cag gca ccg gaa aga gag gaa gcc tgg cgg gta   2251
Arg Val Ile Ser Thr Gln Ala Pro Glu Arg Glu Glu Ala Trp Arg Val
        105                 110                 115

tac ttc gag gct gtt cct gaa ctg gaa gat gat cca cag gca ggc gga   2299
Tyr Phe Glu Ala Val Pro Glu Leu Glu Asp Asp Pro Gln Ala Gly Gly
    120                 125                 130

aag caa aat tca tcc gta agt gtg aat ctt gtc tgg ggg gtg ttg ctg   2347
Lys Gln Asn Ser Ser Val Ser Val Asn Leu Val Trp Gly Val Leu Leu
135                 140                 145                 150

cgt gtt tct ccg tca gac ccc agg cct gcg ctg gta acg gac ggt cac   2395
Arg Val Ser Pro Ser Asp Pro Arg Pro Ala Leu Val Thr Asp Gly His
                155                 160                 165

cac ctg ctg aat acg gga aac aca cgg ctt tct ctt att cgg gct ggc   2443
His Leu Leu Asn Thr Gly Asn Thr Arg Leu Ser Leu Ile Arg Ala Gly
            170                 175                 180

aac tgc gac acc aca tgc cac tgg cag aat ata ggc aaa agt att tat   2491
Asn Cys Asp Thr Thr Cys His Trp Gln Asn Ile Gly Lys Ser Ile Tyr
        185                 190                 195

ccc ggc ggg agt gct gat att ccg gcc gga ata aaa agt aat gca ttt   2539
Pro Gly Gly Ser Ala Asp Ile Pro Ala Gly Ile Lys Ser Asn Ala Phe
    200                 205                 210

cgt gtg gaa tat cgt acg ggt gca aat tca ccg gta atc tct gct gat   2587
Arg Val Glu Tyr Arg Thr Gly Ala Asn Ser Pro Val Ile Ser Ala Asp
215                 220                 225                 230

tta aca gca gcc gga aag taa aaacacacgg agcgtacgct ataccctaca      2638
Leu Thr Ala Ala Gly Lys
                235

tttattctca gggggagcgg atg tat acc gag tgt aca tat atc act gta ata 2691
                        Met Tyr Thr Glu Cys Thr Tyr Ile Thr Val Ile
                                240                 245

aac aac aaa gca agg tta ttt ttt atg aac atg aaa aca tct ttt att   2739
Asn Asn Lys Ala Arg Leu Phe Phe Met Asn Met Lys Thr Ser Phe Ile
```

```
    250                 255                 260

gcc gca gct gtg gca ttg gcc acc gtt tat tct ttt tct gtt tct gcg   2787
Ala Ala Ala Val Ala Leu Ala Thr Val Tyr Ser Phe Ser Val Ser Ala
265                 270                 275                 280

gtt cag aag gat att acc gtc act gcc aat att gac agt aca ctt gaa   2835
Val Gln Lys Asp Ile Thr Val Thr Ala Asn Ile Asp Ser Thr Leu Glu
                285                 290                 295

ctg ctg cag gcc gat ggt tca tcc ctc ccg tcg act atg aag ctg gat   2883
Leu Leu Gln Ala Asp Gly Ser Ser Leu Pro Ser Thr Met Lys Leu Asp
            300                 305                 310

ttc atg ccg ggt aag ggc ctg gtc cat aaa tca ctc cag acc cgc ctt   2931
Phe Met Pro Gly Lys Gly Leu Val His Lys Ser Leu Gln Thr Arg Leu
        315                 320                 325

tac agc aac gat cag acc aag tcg gtt aat gta aaa ctg ttg aat gct   2979
Tyr Ser Asn Asp Gln Thr Lys Ser Val Asn Val Lys Leu Leu Asn Ala
    330                 335                 340

cca caa ctt atc aac gtc ctg gat ccc acc aaa acc att gat atg gaa   3027
Pro Gln Leu Ile Asn Val Leu Asp Pro Thr Lys Thr Ile Asp Met Glu
345                 350                 355                 360

gtg act ctg gga gga cgg tca ctg acc acc acc aat tct gta ctg gaa   3075
Val Thr Leu Gly Gly Arg Ser Leu Thr Thr Thr Asn Ser Val Leu Glu
                365                 370                 375

gct aaa acc ctg ttc ccg gac gga aaa act ggc gat gct tca gct ctg   3123
Ala Lys Thr Leu Phe Pro Asp Gly Lys Thr Gly Asp Ala Ser Ala Leu
            380                 385                 390

ctg aac ctg gat att ggt cag aag gct gga gca gcc tta caa aac ctg   3171
Leu Asn Leu Asp Ile Gly Gln Lys Ala Gly Ala Ala Leu Gln Asn Leu
        395                 400                 405

cct gcc ggt gaa tac agc gga ttg gtc agt ctg gtg att tca cag gct   3219
Pro Ala Gly Glu Tyr Ser Gly Leu Val Ser Leu Val Ile Ser Gln Ala
    410                 415                 420

gtc act gcc ggc taa taactggtta ttagctcttc atctgatccg gttttggggg   3274
Val Thr Ala Gly
425

gcaccgttcg tacctgaacc ggatccggta ttgatcttat tattcattgc aattcaggtc 3334

tctttacgtg agtcgttatt tctgg atg tat tat tta ctg gga ttg tgc agt   3386
                           Met Tyr Tyr Leu Leu Gly Leu Cys Ser
                           430                 435

ttt acc agc cag gca act ctt att ccc cct cct gga ttt gaa tct ctg   3434
Phe Thr Ser Gln Ala Thr Leu Ile Pro Pro Pro Gly Phe Glu Ser Leu
    440                 445                 450

ctg gaa gga cag act gag caa att gaa gtg ttg cta cca ggg cat tca   3482
Leu Glu Gly Gln Thr Glu Gln Ile Glu Val Leu Leu Pro Gly His Ser
455                 460                 465                 470
```

```
ctg gga tta ttt ccg gtg gtg gtt aaa ccg gac acc gtg cag ttc atg   3530
Leu Gly Leu Phe Pro Val Val Val Lys Pro Asp Thr Val Gln Phe Met
                475                 480                 485

tcc cca ttg atg gta ctt gaa agc agt ggg ctt gcc gcg ttg ccg gcc   3578
Ser Pro Leu Met Val Leu Glu Ser Ser Gly Leu Ala Ala Leu Pro Ala
            490                 495                 500

gca gaa cgg caa aaa gcg ctg gct gca ctc agc cgt ccg ttg cta cgt   3626
Ala Glu Arg Gln Lys Ala Leu Ala Ala Leu Ser Arg Pro Leu Leu Arg
        505                 510                 515

aac agc aat ctg gtc tgt ggt gtc tca gaa gca aaa gac agc agc gag   3674
Asn Ser Asn Leu Val Cys Gly Val Ser Glu Ala Lys Asp Ser Ser Glu
    520                 525                 530

tgt ggt tac gtg gca aca gat aaa gag gat gtt gcg gtt att ttt gat   3722
Cys Gly Tyr Val Ala Thr Asp Lys Glu Asp Val Ala Val Ile Phe Asp
535                 540                 545                 550

gag aac aac gct cag tta tct ttg ttt ctt aac cgg gac tgg ttg ccg   3770
Glu Asn Asn Ala Gln Leu Ser Leu Phe Leu Asn Arg Asp Trp Leu Pro
                555                 560                 565

gat gaa gaa cga cgt gat aaa cgc tgg ctg act ccg acc ccg gag ggt   3818
Asp Glu Glu Arg Arg Asp Lys Arg Trp Leu Thr Pro Thr Pro Glu Gly
            570                 575                 580

gtc agc gca ttt att cac cgc cag acg ctg tat ctg agt gat gat ctc   3866
Val Ser Ala Phe Ile His Arg Gln Thr Leu Tyr Leu Ser Asp Asp Leu
        585                 590                 595

cac agt cgt aat atg aca ctg aat ggt agc ggt gcc ctg ggg ctt ggt   3914
His Ser Arg Asn Met Thr Leu Asn Gly Ser Gly Ala Leu Gly Leu Gly
    600                 605                 610

gac ggt cgt tat ctg gga ggc gac tgg gcg gct atc tgg aat cag tca   3962
Asp Gly Arg Tyr Leu Gly Gly Asp Trp Ala Ala Ile Trp Asn Gln Ser
615                 620                 625                 630

gaa cat tac aat aac agt cag gcc tgg ttt gac aat ctg ttt gtc cgt   4010
Glu His Tyr Asn Asn Ser Gln Ala Trp Phe Asp Asn Leu Phe Val Arg
                635                 640                 645

cag gat ctc ggc aat cag tat tat ctc cag gct ggt cgg atg gat cag   4058
Gln Asp Leu Gly Asn Gln Tyr Tyr Leu Gln Ala Gly Arg Met Asp Gln
            650                 655                 660

cgg aat ctg tcc agc gcc acg ggg ggg gat ttt ggg ttc agt ctg ctt   4106
Arg Asn Leu Ser Ser Ala Thr Gly Gly Asp Phe Gly Phe Ser Leu Leu
        665                 670                 675

ccc ctg agc cgg ttt gat gga tta cga acc ggg acc acc caa gct tat   4154
Pro Leu Ser Arg Phe Asp Gly Leu Arg Thr Gly Thr Thr Gln Ala Tyr
    680                 685                 690

gtt aac cat gag gtg gac cat aat gcc act ccg gtt atg gtt cag gtt   4202
Val Asn His Glu Val Asp His Asn Ala Thr Pro Val Met Val Gln Val
695                 700                 705                 710
```

```
acc cga aat gcc cgt att gat att tat cgt ggc agc gag ttg ctg ggg   4250
Thr Arg Asn Ala Arg Ile Asp Ile Tyr Arg Gly Ser Glu Leu Leu Gly
                715                 720                 725

agt cag ttc ctg acc ccg gga atg cat acc ctg gat act cat tct ctt   4298
Ser Gln Phe Leu Thr Pro Gly Met His Thr Leu Asp Thr His Ser Leu
            730                 735                 740

cca ccg gga agc tat cct ctg gcg ttg cgg gtg tat gag gat ggg att   4346
Pro Pro Gly Ser Tyr Pro Leu Ala Leu Arg Val Tyr Glu Asp Gly Ile
        745                 750                 755

ctg cgg cga acg gag acc cag ccc ttc agt aag ggg ggc aat agc ttc   4394
Leu Arg Arg Thr Glu Thr Gln Pro Phe Ser Lys Gly Gly Asn Ser Phe
    760                 765                 770

agt gca cag acc cag tgg ttt att cag ggc ggg ctg gaa gat acc ggg   4442
Ser Ala Gln Thr Gln Trp Phe Ile Gln Gly Gly Leu Glu Asp Thr Gly
775                 780                 785                 790

gat aaa gcc agc cat tat gac ggt gag act gtc atg gct gcc gga ttc   4490
Asp Lys Ala Ser His Tyr Asp Gly Glu Thr Val Met Ala Ala Gly Phe
                795                 800                 805

caa act ggg ctg cgg aaa aat atc agt ctg acc gaa ggt atc tct ctg   4538
Gln Thr Gly Leu Arg Lys Asn Ile Ser Leu Thr Glu Gly Ile Ser Leu
            810                 815                 820

gca cat gag gcc tgg tac agt gaa acc cga ctg aat tca cag cat gca   4586
Ala His Glu Ala Trp Tyr Ser Glu Thr Arg Leu Asn Ser Gln His Ala
        825                 830                 835

gtg ctg gat ggc acg ctg gac ctt tct gcc ggg ata ctg cat ggg aca   4634
Val Leu Asp Gly Thr Leu Asp Leu Ser Ala Gly Ile Leu His Gly Thr
    840                 845                 850

gac agc acg agc ggt aac act gag cag gtg aca tac aac gac gga ttt   4682
Asp Ser Thr Ser Gly Asn Thr Glu Gln Val Thr Tyr Asn Asp Gly Phe
855                 860                 865                 870

tcc gcg agt ctg tgg cgt aac cat acg gaa agt gat gcc tgt agt ggt   4730
Ser Ala Ser Leu Trp Arg Asn His Thr Glu Ser Asp Ala Cys Ser Gly
                875                 880                 885

cgt cat cca cag tca gtg cat gcc agt atg acc tgc cag act tcg atg   4778
Arg His Pro Gln Ser Val His Ala Ser Met Thr Cys Gln Thr Ser Met
            890                 895                 900

aac gcc tcc ctg tcg gtt tcg gtg ggg aac tgg tat gcc cta ctg gga   4826
Asn Ala Ser Leu Ser Val Ser Val Gly Asn Trp Tyr Ala Leu Leu Gly
        905                 910                 915

tac agt acc agc agg aca gaa ggt cgg ccg gtt tac cgg gga tat gat   4874
Tyr Ser Thr Ser Arg Thr Glu Gly Arg Pro Val Tyr Arg Gly Tyr Asp
    920                 925                 930

gat aac agt gac aaa gaa aat gtg ttc tgg cga cag gca tac atc cct   4922
Asp Asn Ser Asp Lys Glu Asn Val Phe Trp Arg Gln Ala Tyr Ile Pro
935                 940                 945                 950
```

```
gcc tct cac cgc gaa tct gct cag gct agt gca acg tac agc ctt aat   4970
Ala Ser His Arg Glu Ser Ala Gln Ala Ser Ala Thr Tyr Ser Leu Asn
                955                 960                 965

atg gct ggc atg aat att aat acc cat ggg gga gta tgg cga acc cga   5018
Met Ala Gly Met Asn Ile Asn Thr His Gly Gly Val Trp Arg Thr Arg
            970                 975                 980

aat gac gga gtg aat gat gat ggc ttg ttt atg agt gtc agt gtg tca   5066
Asn Asp Gly Val Asn Asp Asp Gly Leu Phe Met Ser Val Ser Val Ser
        985                 990                 995

tat gcc tct caa cca ccg aca atg act ggc agt aat agg tat acc tca   5114
Tyr Ala Ser Gln Pro Pro Thr Met Thr Gly Ser Asn Arg Tyr Thr Ser
   1000                1005                1010


gcc ggg acc gat att cac agt agc cgg aat caa aaa aca cag acg tcc   5162
Ala Gly Thr Asp Ile His Ser Ser Arg Asn Gln Lys Thr Gln Thr Ser
1015                1020                1025                1030

tgg aat gtg aac cat gtg aga tcc tgg cag cag gat ctg tat cgt gaa   5210
Trp Asn Val Asn His Val Arg Ser Trp Gln Gln Asp Leu Tyr Arg Glu
               1035                1040                1045

ctg tcg gtg ggt ttc tcc ggt tat aac gac gac agc tgg agc ggg agt   5258
Leu Ser Val Gly Phe Ser Gly Tyr Asn Asp Asp Ser Trp Ser Gly Ser
           1050                1055                1060

ctc ggc gga cgc atg agc ggc cgt atg ggt gaa ctg agc gcc act atc   5306
Leu Gly Gly Arg Met Ser Gly Arg Met Gly Glu Leu Ser Ala Thr Ile
       1065                1070                1075

agt aac tcc cat caa cgt aat gcg ggc agc gcc agt tca ctc acc gct   5354
Ser Asn Ser His Gln Arg Asn Ala Gly Ser Ala Ser Ser Leu Thr Ala
   1080                1085                1090

ggc tac agc tcg tct ctg gcg tta tcc cgt aat gga ctg ttc tgg gga   5402
Gly Tyr Ser Ser Ser Leu Ala Leu Ser Arg Asn Gly Leu Phe Trp Gly
1095                1100                1105                1110

ggt ggt cag gac ggt gaa ccg gcc tct ggc atg gcg gtg aac gtg gag   5450
Gly Gly Gln Asp Gly Glu Pro Ala Ser Gly Met Ala Val Asn Val Glu
               1115                1120                1125

tca gag ggg gac gag ggc agt agc ggg aaa gta gtc agc gtt cgt ggc   5498
Ser Glu Gly Asp Glu Gly Ser Ser Gly Lys Val Val Ser Val Arg Gly
           1130                1135                1140

agc agc cag ccg ttc agt ctc ggt ttt ggt cag cag tcg ctg ttg ctg   5546
Ser Ser Gln Pro Phe Ser Leu Gly Phe Gly Gln Gln Ser Leu Leu Leu
       1145                1150                1155

atg gaa ggc tat aac gcc acg gag gtg acc att gag gat gca ggg gtt   5594
Met Glu Gly Tyr Asn Ala Thr Glu Val Thr Ile Glu Asp Ala Gly Val
   1160                1165                1170

agt tca cag ggt atg gca ggc gta aaa gcg gga ggg gga agc agg tgt   5642
Ser Ser Gln Gly Met Ala Gly Val Lys Ala Gly Gly Gly Ser Arg Cys
```

```
1175                1180                1185                1190

tac ttc ctg aca ccc ggg cat ctg ctg gtt cac aac atc agc gcc agt   5690
Tyr Phe Leu Thr Pro Gly His Leu Leu Val His Asn Ile Ser Ala Ser
                1195                1200                1205

atg agc cga ctg tac gtt ggc cgc gta ctg gac aag gat ggc aga ccg   5738
Met Ser Arg Leu Tyr Val Gly Arg Val Leu Asp Lys Asp Gly Arg Pro
            1210                1215                1220

ctg ctg gac gca cag cca ctg aac tat cca ttt ttg tcg ttg gga cct   5786
Leu Leu Asp Ala Gln Pro Leu Asn Tyr Pro Phe Leu Ser Leu Gly Pro
        1225                1230                1235

tcc ggg cga ttt agc ctg cag agc gag cat aaa gaa tcc agc ctg tgg   5834
Ser Gly Arg Phe Ser Leu Gln Ser Glu His Lys Glu Ser Ser Leu Trp
    1240                1245                1250

ctg ctg tct aaa aac agg atc ctg cgt tgt ccg atg tca gta cat aaa   5882
Leu Leu Ser Lys Asn Arg Ile Leu Arg Cys Pro Met Ser Val His Lys
1255                1260                1265                1270

cgt cgg gat gtt atg cag gta gtg ggt gat gtg cgg tgt gaa tta agt   5930
Arg Arg Asp Val Met Gln Val Val Gly Asp Val Arg Cys Glu Leu Ser
                1275                1280                1285

gac gtg gat gcc ctg cca cag gcg ttg caa ata tcg ccg cgg gtc atc   5978
Asp Val Asp Ala Leu Pro Gln Ala Leu Gln Ile Ser Pro Arg Val Ile
            1290                1295                1300

cgt ttg ctg aac gtg gca ggt ttg ctg cgc cat tcc gtt cag gaa gcc   6026
Arg Leu Leu Asn Val Ala Gly Leu Leu Arg His Ser Val Gln Glu Ala
        1305                1310                1315

tga cgtagagata aaggcgttaa ct atg agt aat aaa atg aag tgg acg agt  6078
                            Met Ser Asn Lys Met Lys Trp Thr Ser
                        1320                1325

atg aca gcc cat tgg tca gca att att aat ttc atc cga aaa tat gtt   6126
Met Thr Ala His Trp Ser Ala Ile Ile Asn Phe Ile Arg Lys Tyr Val
    1330                1335                1340

tat cca gca agg ata att gcc atc ctg ctg atg gct ggc gct aca ctg   6174
Tyr Pro Ala Arg Ile Ile Ala Ile Leu Leu Met Ala Gly Ala Thr Leu
1345                1350                1355                1360

cca caa gtc gcc gat gcg att acc gtc gac ctg aat tac gac aag aac   6222
Pro Gln Val Ala Asp Ala Ile Thr Val Asp Leu Asn Tyr Asp Lys Asn
                1365                1370                1375

aat gta gcg gtc atc act cct gtc tgg tcc caa gaa tgg agt gta gca   6270
Asn Val Ala Val Ile Thr Pro Val Trp Ser Gln Glu Trp Ser Val Ala
            1380                1385                1390

aat gtg ttg ggg gga tgg gta tgt cgt tca aac agg aat gaa aat gag   6318
Asn Val Leu Gly Gly Trp Val Cys Arg Ser Asn Arg Asn Glu Asn Glu
        1395                1400                1405

ggg gcg tgt gaa gaa aca cat ttg gta tgg tgg tat gct ttt gga gct   6366
Gly Ala Cys Glu Glu Thr His Leu Val Trp Trp Tyr Ala Phe Gly Ala
```

```
   1410                1415                1420

tat tca aaa att cgt ctg cgt ttc aga gaa caa atc agc cat gcc gaa   6414
Tyr Ser Lys Ile Arg Leu Arg Phe Arg Glu Gln Ile Ser His Ala Glu
1425                1430                1435                1440

att acg ctc ata ctg ctc ggc agt gtt cgt gat gcc tgt tat act ggt   6462
Ile Thr Leu Ile Leu Leu Gly Ser Val Arg Asp Ala Cys Tyr Thr Gly
                1445                1450                1455

gtc atc aac atg aac gct gct gca tgt caa tgg ggt agg tcg ctg aaa   6510
Val Ile Asn Met Asn Ala Ala Ala Cys Gln Trp Gly Arg Ser Leu Lys
            1460                1465                1470

ctt agg ata cct tca gaa gag ctt gcg aag ata cct aca agc gga aca   6558
Leu Arg Ile Pro Ser Glu Glu Leu Ala Lys Ile Pro Thr Ser Gly Thr
        1475                1480                1485

tgg aaa gca acg tta gtc ctg gat tat tta caa tgg ggc gga gac gat   6606
Trp Lys Ala Thr Leu Val Leu Asp Tyr Leu Gln Trp Gly Gly Asp Asp
    1490                1495                1500



cct tta ggc aca tca act aca gat atc acg ctg aat gta aca gac cac   6654
Pro Leu Gly Thr Ser Thr Thr Asp Ile Thr Leu Asn Val Thr Asp His
1505                1510                1515                1520

ttt gct gaa aat gcg gct att tac ttt ccg caa ttt ggt aca gca acg   6702
Phe Ala Glu Asn Ala Ala Ile Tyr Phe Pro Gln Phe Gly Thr Ala Thr
                1525                1530                1535

ccc cgg gtg gac ctg aat ctt cac cgg atg aat gcc tca caa atg tcg   6750
Pro Arg Val Asp Leu Asn Leu His Arg Met Asn Ala Ser Gln Met Ser
            1540                1545                1550

ggc agg gct aat ctg gat atg tgt ctg tat gac gga ggt gtg aaa gcc   6798
Gly Arg Ala Asn Leu Asp Met Cys Leu Tyr Asp Gly Gly Val Lys Ala
        1555                1560                1565

cgt tca tta cag atg aag ata gaa gga agc aat aag tca ggt acg gga   6846
Arg Ser Leu Gln Met Lys Ile Glu Gly Ser Asn Lys Ser Gly Thr Gly
    1570                1575                1580

ttt cag gtt ata aag agc gat tct gct gat acg att gat tat gcg gtc   6894
Phe Gln Val Ile Lys Ser Asp Ser Ala Asp Thr Ile Asp Tyr Ala Val
1585                1590                1595                1600

agt atg aat tat ggg gga cga agt att cct gtc acc cgt ggc gtg gag   6942
Ser Met Asn Tyr Gly Gly Arg Ser Ile Pro Val Thr Arg Gly Val Glu
                1605                1610                1615

ttc agt ctg gat aac gtg gat aaa gca gca acg cgt ccg gtg gta ctt   6990
Phe Ser Leu Asp Asn Val Asp Lys Ala Ala Thr Arg Pro Val Val Leu
            1620                1625                1630

ccc ggg caa cgg cag gcg gta cgt tgt gtg cca gtg ccc ctt acc ctg   7038
Pro Gly Gln Arg Gln Ala Val Arg Cys Val Pro Val Pro Leu Thr Leu
        1635                1640                1645
```

```
aca aca caa ccc ttt aac atc aga gag aag cgt tct ggt gag tat cag   7086
Thr Thr Gln Pro Phe Asn Ile Arg Glu Lys Arg Ser Gly Glu Tyr Gln
   1650                1655                1660

gga acg ctg aca gtg aca atg ctg atg gga aca caa acc ccc tga       7131
Gly Thr Leu Thr Val Thr Met Leu Met Gly Thr Gln Thr Pro
1665             1670                1675

cagtaattat ttatttttatt gatatctttc ttatatggtt ttttaaatca gagttctctt 7191

tatatacttg ttttatttaa taaagagaat ctattcactt atgaaaatca atgcgtgagg 7251

ttctgctttc ct atg act gtg tat tta gat gat aaa gat aaa gaa tta ttg 7302
              Met Thr Val Tyr Leu Asp Asp Lys Asp Lys Glu Leu Leu
            1680                1685                1690

aaa gaa atc caa aaa gat tgt gca caa act tta tgg caa ctt gca tat   7350
Lys Glu Ile Gln Lys Asp Cys Ala Gln Thr Leu Trp Gln Leu Ala Tyr
       1695                1700                1705

aaa gtg gga ctt acg ccc aca cca tgt ttc aaa cgt tta aaa aaa ctt   7398
Lys Val Gly Leu Thr Pro Thr Pro Cys Phe Lys Arg Leu Lys Lys Leu
   1710                1715                1720

aaa gac agg ggg gtt atc att ggt cag ttc gct tta ttg gat aag gaa   7446
Lys Asp Arg Gly Val Ile Ile Gly Gln Phe Ala Leu Leu Asp Lys Glu
1725                1730                1735                1740

aaa cta ggt ctt tca ctt aat gtc ttt att atg att aac ata tct gag   7494
Lys Leu Gly Leu Ser Leu Asn Val Phe Ile Met Ile Asn Ile Ser Glu
               1745                1750                1755

gag caa tac gct agt att tct gag aaa ata aag tca atg cct gag gtt   7542
Glu Gln Tyr Ala Ser Ile Ser Glu Lys Ile Lys Ser Met Pro Glu Val
           1760                1765                1770

att gcc ttt tat cga att tct gga tca ttt aat tat tta atg cat aca   7590
Ile Ala Phe Tyr Arg Ile Ser Gly Ser Phe Asn Tyr Leu Met His Thr
       1775                1780                1785

gta ttt aca gat atg aac gat tac tat agt ttt tat gag aaa ata ata   7638
Val Phe Thr Asp Met Asn Asp Tyr Tyr Ser Phe Tyr Glu Lys Ile Ile
   1790                1795                1800

tta act aat tct tca att agt gga tct gca tcg agc ttt gtt ctt gag   7686
Leu Thr Asn Ser Ser Ile Ser Gly Ser Ala Ser Ser Phe Val Leu Glu
1805                1810                1815                1820

caa ata aag gaa aca aac gaa ctg tca gtg tga aagtgtgatg tgtacttact 7739
Gln Ile Lys Glu Thr Asn Glu Leu Ser Val
               1825                1830

gatttaatac attattatcc ttcttacgga acaacaacgg cagattgcgg ctgttgaaca 7799

aggattttaa tcagcagtgg tgaaattaag cggcacagaa taacacagcg gaatatcaca 7859

tggttaaata tcaccccgtg catgtaacaa aaaaccgcat taaaacagat gatgttactg 7919

atatttattt cgttgaaccc ttctggaaaa aaggcgaaaa ccacataatt gagtcattga 7979
```

```
tgtttttga agagttacaa aagtcattta atttattcaa ccataaatat gggttaaata 8039

aatatatact caggatcccc tgggaatttg tgctcataca tatggaaagg atcagtaaat 8099

taaatagcgt cgggttattt gctgtttctg ttgactttaa taacaaccac aaatttctga 8159

gcgagtacat caggagtcgc agagattatg gtatggaagt ttggtttgat ttttgtggta 8219

aacattctta ttccagtgaa attaaaaacc ttggattctt ttttcaggct tgcgtagtgc 8279

ctcgtgatcc taatttttatt agtagtgttt atcattatca taagttccaa aagattcttg 8339

tcggggatat aaatgatgta gaacagaggg ccgtgtacca gaacgaagtt gattacatgt 8399

atggaatgca atggccatcg tcatatgacg gttttttctt tcgggatcat aaaaaaaatg 8459

aaacttggtg tatataacag aaggagtgaa aatttgaatc aaaaatatct tatttatttt 8519

ttgtttaatt attgttttgt tttttattac gattaaatat aaagaacatc attgttcgtg 8579

cggtggggag gctggaagtt taggggatga ccgtttatca acaattttat tacagccacc 8639

atacgaatgg tttatatatg cactagatgt attattttag tttaatatat cgatggttgc 8699

tatttgcatt gatgatgttc cgttacatta aggaatatac atctgtatct cgttatacgc 8759

acactcacat tactaatcat tattaatatg agtgtggttc ttgttttacg catgcatggt 8819

tgcatgtgac gttaaattta aatgagctga ctgtatgaat tctaaatact ttagagaggt 8879

gttttttgtc tcggtagttg ttatattatt attttatttg gtgttatttg cagccagtgc 8939

tcatgctgaa ggcggtttca gatctggagg cattgggtta tttatgacgg gaacaagaga 8999

gatgctactg tagagataat aaattctgct aaagattccc caattcttgt gcattgacat 9059

cctccacgtc ctgaagggcg tgggttcctg ctccaacggg ctgcctgact gcacgctcct 9119

tccacaggca agcacggcgt gtcccgctct aaaatgttac gcgcgccgtt tacatcggcg 9179

ttcgcagtat atcttcatac cagacacttg taagtatctc gcataatcgt gccattcaca 9239

tttagagatc atac                                                   9253
```

<210> 2
<211> 236
<212> PRT
<213> Salmonella typhi

<400> 2

```
Met Asn Phe Lys Asp Thr Leu Pro Gly Val Phe Leu Cys Val Ala Met
  1               5                  10                  15

Phe Ala Cys Gly His Ala Arg Ala Asn Met Leu Val Tyr Pro Met Ala
             20                  25                  30

Ala Glu Ile Asn Ser Ser Arg Glu Glu Ala Thr Ser Leu Phe Val Tyr
         35                  40                  45
```

```
Ser Lys Ser Asp His Val Gln Tyr Ile Arg Thr Arg Ile Met Arg Ile
     50                  55                  60

Glu His Pro Gly Met Pro Gln Glu Lys Glu Val Pro Ala Gly Asn Asp
 65                  70                  75                  80

Ile Glu Thr Gly Leu Val Val Ser Pro Glu Lys Phe Ala Leu Ser Pro
                 85                  90                  95

Gly Thr Lys Lys Thr Ile Arg Val Ile Ser Thr Gln Ala Pro Glu Arg
            100                 105                 110

Glu Glu Ala Trp Arg Val Tyr Phe Glu Ala Val Pro Glu Leu Glu Asp
        115                 120                 125

Asp Pro Gln Ala Gly Gly Lys Gln Asn Ser Ser Val Ser Val Asn Leu
    130                 135                 140

Val Trp Gly Val Leu Leu Arg Val Ser Pro Ser Asp Pro Arg Pro Ala
145                 150                 155                 160

Leu Val Thr Asp Gly His His Leu Leu Asn Thr Gly Asn Thr Arg Leu
                165                 170                 175

Ser Leu Ile Arg Ala Gly Asn Cys Asp Thr Thr Cys His Trp Gln Asn
            180                 185                 190

Ile Gly Lys Ser Ile Tyr Pro Gly Gly Ser Ala Asp Ile Pro Ala Gly
        195                 200                 205

Ile Lys Ser Asn Ala Phe Arg Val Glu Tyr Arg Thr Gly Ala Asn Ser
    210                 215                 220

Pro Val Ile Ser Ala Asp Leu Thr Ala Ala Gly Lys
225                 230                 235


<210> 3
<211> 191
<212> PRT
<213> Salmonella typhi

<400> 3
Met Tyr Thr Glu Cys Thr Tyr Ile Thr Val Ile Asn Asn Lys Ala Arg
  1               5                  10                  15

Leu Phe Phe Met Asn Met Lys Thr Ser Phe Ile Ala Ala Ala Val Ala
             20                  25                  30

Leu Ala Thr Val Tyr Ser Phe Ser Val Ser Ala Val Gln Lys Asp Ile
         35                  40                  45

Thr Val Thr Ala Asn Ile Asp Ser Thr Leu Glu Leu Leu Gln Ala Asp
     50                  55                  60

Gly Ser Ser Leu Pro Ser Thr Met Lys Leu Asp Phe Met Pro Gly Lys
 65                  70                  75                  80

Gly Leu Val His Lys Ser Leu Gln Thr Arg Leu Tyr Ser Asn Asp Gln
                 85                  90                  95
```

```
Thr Lys Ser Val Asn Val Lys Leu Leu Asn Ala Pro Gln Leu Ile Asn
            100                 105                 110

Val Leu Asp Pro Thr Lys Thr Ile Asp Met Glu Val Thr Leu Gly Gly
        115                 120                 125

Arg Ser Leu Thr Thr Thr Asn Ser Val Leu Glu Ala Lys Thr Leu Phe
    130                 135                 140

Pro Asp Gly Lys Thr Gly Asp Ala Ser Ala Leu Leu Asn Leu Asp Ile
145                 150                 155                 160

Gly Gln Lys Ala Gly Ala Ala Leu Gln Asn Leu Pro Ala Gly Glu Tyr
                165                 170                 175

Ser Gly Leu Val Ser Leu Val Ile Ser Gln Ala Val Thr Ala Gly
            180                 185                 190


<210> 4
<211> 889
<212> PRT
<213> Salmonella typhi

<400> 4
Met Tyr Tyr Leu Leu Gly Leu Cys Ser Phe Thr Ser Gln Ala Thr Leu
  1               5                  10                  15

Ile Pro Pro Pro Gly Phe Glu Ser Leu Leu Glu Gly Gln Thr Glu Gln
             20                  25                  30

Ile Glu Val Leu Leu Pro Gly His Ser Leu Gly Leu Phe Pro Val Val
         35                  40                  45


Val Lys Pro Asp Thr Val Gln Phe Met Ser Pro Leu Met Val Leu Glu
     50                  55                  60

Ser Ser Gly Leu Ala Ala Leu Pro Ala Ala Glu Arg Gln Lys Ala Leu
 65                  70                  75                  80

Ala Ala Leu Ser Arg Pro Leu Leu Arg Asn Ser Asn Leu Val Cys Gly
                85                  90                  95

Val Ser Glu Ala Lys Asp Ser Ser Glu Cys Gly Tyr Val Ala Thr Asp
            100                 105                 110

Lys Glu Asp Val Ala Val Ile Phe Asp Glu Asn Asn Ala Gln Leu Ser
        115                 120                 125

Leu Phe Leu Asn Arg Asp Trp Leu Pro Asp Glu Glu Arg Arg Asp Lys
    130                 135                 140

Arg Trp Leu Thr Pro Thr Pro Glu Gly Val Ser Ala Phe Ile His Arg
145                 150                 155                 160

Gln Thr Leu Tyr Leu Ser Asp Asp Leu His Ser Arg Asn Met Thr Leu
                165                 170                 175
```

```
Asn Gly Ser Gly Ala Leu Gly Leu Gly Asp Gly Arg Tyr Leu Gly Gly
            180                 185                 190

Asp Trp Ala Ala Ile Trp Asn Gln Ser Glu His Tyr Asn Asn Ser Gln
        195                 200                 205

Ala Trp Phe Asp Asn Leu Phe Val Arg Gln Asp Leu Gly Asn Gln Tyr
    210                 215                 220

Tyr Leu Gln Ala Gly Arg Met Asp Gln Arg Asn Leu Ser Ser Ala Thr
225                 230                 235                 240

Gly Gly Asp Phe Gly Phe Ser Leu Leu Pro Leu Ser Arg Phe Asp Gly
                245                 250                 255

Leu Arg Thr Gly Thr Thr Gln Ala Tyr Val Asn His Glu Val Asp His
            260                 265                 270

Asn Ala Thr Pro Val Met Val Gln Val Thr Arg Asn Ala Arg Ile Asp
        275                 280                 285

Ile Tyr Arg Gly Ser Glu Leu Leu Gly Ser Gln Phe Leu Thr Pro Gly
    290                 295                 300

Met His Thr Leu Asp Thr His Ser Leu Pro Pro Gly Ser Tyr Pro Leu
305                 310                 315                 320

Ala Leu Arg Val Tyr Glu Asp Gly Ile Leu Arg Arg Thr Glu Thr Gln
                325                 330                 335

Pro Phe Ser Lys Gly Gly Asn Ser Phe Ser Ala Gln Thr Gln Trp Phe
            340                 345                 350

Ile Gln Gly Gly Leu Glu Asp Thr Gly Asp Lys Ala Ser His Tyr Asp
        355                 360                 365

Gly Glu Thr Val Met Ala Ala Gly Phe Gln Thr Gly Leu Arg Lys Asn
    370                 375                 380

Ile Ser Leu Thr Glu Gly Ile Ser Leu Ala His Glu Ala Trp Tyr Ser
385                 390                 395                 400

Glu Thr Arg Leu Asn Ser Gln His Ala Val Leu Asp Gly Thr Leu Asp
                405                 410                 415

Leu Ser Ala Gly Ile Leu His Gly Thr Asp Ser Thr Ser Gly Asn Thr
            420                 425                 430

Glu Gln Val Thr Tyr Asn Asp Gly Phe Ser Ala Ser Leu Trp Arg Asn
        435                 440                 445

His Thr Glu Ser Asp Ala Cys Ser Gly Arg His Pro Gln Ser Val His
    450                 455                 460

Ala Ser Met Thr Cys Gln Thr Ser Met Asn Ala Ser Leu Ser Val Ser
465                 470                 475                 480

Val Gly Asn Trp Tyr Ala Leu Leu Gly Tyr Ser Thr Ser Arg Thr Glu
                485                 490                 495
```

```
Gly Arg Pro Val Tyr Arg Gly Tyr Asp Asp Asn Ser Asp Lys Glu Asn
            500                 505                 510

Val Phe Trp Arg Gln Ala Tyr Ile Pro Ala Ser His Arg Glu Ser Ala
        515                 520                 525

Gln Ala Ser Ala Thr Tyr Ser Leu Asn Met Ala Gly Met Asn Ile Asn
    530                 535                 540

Thr His Gly Gly Val Trp Arg Thr Arg Asn Asp Gly Val Asn Asp Asp
545                 550                 555                 560

Gly Leu Phe Met Ser Val Ser Val Ser Tyr Ala Ser Gln Pro Pro Thr
                565                 570                 575

Met Thr Gly Ser Asn Arg Tyr Thr Ser Ala Gly Thr Asp Ile His Ser
            580                 585                 590

Ser Arg Asn Gln Lys Thr Gln Thr Ser Trp Asn Val Asn His Val Arg
        595                 600                 605

Ser Trp Gln Gln Asp Leu Tyr Arg Glu Leu Ser Val Gly Phe Ser Gly
    610                 615                 620

Tyr Asn Asp Asp Ser Trp Ser Gly Ser Leu Gly Gly Arg Met Ser Gly
625                 630                 635                 640

Arg Met Gly Glu Leu Ser Ala Thr Ile Ser Asn Ser His Gln Arg Asn
                645                 650                 655

Ala Gly Ser Ala Ser Ser Leu Thr Ala Gly Tyr Ser Ser Ser Leu Ala
            660                 665                 670

Leu Ser Arg Asn Gly Leu Phe Trp Gly Gly Gly Gln Asp Gly Glu Pro
        675                 680                 685

Ala Ser Gly Met Ala Val Asn Val Glu Ser Glu Gly Asp Glu Gly Ser
    690                 695                 700

Ser Gly Lys Val Val Ser Val Arg Gly Ser Ser Gln Pro Phe Ser Leu
705                 710                 715                 720

Gly Phe Gly Gln Gln Ser Leu Leu Leu Met Glu Gly Tyr Asn Ala Thr
                725                 730                 735

Glu Val Thr Ile Glu Asp Ala Gly Val Ser Ser Gln Gly Met Ala Gly
            740                 745                 750

Val Lys Ala Gly Gly Gly Ser Arg Cys Tyr Phe Leu Thr Pro Gly His
        755                 760                 765

Leu Leu Val His Asn Ile Ser Ala Ser Met Ser Arg Leu Tyr Val Gly
    770                 775                 780

Arg Val Leu Asp Lys Asp Gly Arg Pro Leu Leu Asp Ala Gln Pro Leu
785                 790                 795                 800

Asn Tyr Pro Phe Leu Ser Leu Gly Pro Ser Gly Arg Phe Ser Leu Gln
                805                 810                 815
```

```
Ser Glu His Lys Glu Ser Ser Leu Trp Leu Leu Ser Lys Asn Arg Ile
            820                 825                 830

Leu Arg Cys Pro Met Ser Val His Lys Arg Arg Asp Val Met Gln Val
        835                 840                 845

Val Gly Asp Val Arg Cys Glu Leu Ser Asp Val Asp Ala Leu Pro Gln
    850                 855                 860

Ala Leu Gln Ile Ser Pro Arg Val Ile Arg Leu Leu Asn Val Ala Gly
865                 870                 875                 880

Leu Leu Arg His Ser Val Gln Glu Ala
                885
```

<210> 5
<211> 359
<212> PRT
<213> Salmonella typhi

```
<400> 5
Met Ser Asn Lys Met Lys Trp Thr Ser Met Thr Ala His Trp Ser Ala
  1               5                  10                  15

Ile Ile Asn Phe Ile Arg Lys Tyr Val Tyr Pro Ala Arg Ile Ile Ala
             20                  25                  30

Ile Leu Leu Met Ala Gly Ala Thr Leu Pro Gln Val Ala Asp Ala Ile
         35                  40                  45

Thr Val Asp Leu Asn Tyr Asp Lys Asn Asn Val Ala Val Ile Thr Pro
     50                  55                  60

Val Trp Ser Gln Glu Trp Ser Val Ala Asn Val Leu Gly Gly Trp Val
 65                  70                  75                  80

Cys Arg Ser Asn Arg Asn Glu Asn Glu Gly Ala Cys Glu Glu Thr His
                85                  90                  95

Leu Val Trp Trp Tyr Ala Phe Gly Ala Tyr Ser Lys Ile Arg Leu Arg
            100                 105                 110

Phe Arg Glu Gln Ile Ser His Ala Glu Ile Thr Leu Ile Leu Leu Gly
        115                 120                 125

Ser Val Arg Asp Ala Cys Tyr Thr Gly Val Ile Asn Met Asn Ala Ala
    130                 135                 140

Ala Cys Gln Trp Gly Arg Ser Leu Lys Leu Arg Ile Pro Ser Glu Glu
145                 150                 155                 160

Leu Ala Lys Ile Pro Thr Ser Gly Thr Trp Lys Ala Thr Leu Val Leu
                165                 170                 175

Asp Tyr Leu Gln Trp Gly Gly Asp Asp Pro Leu Gly Thr Ser Thr Thr
            180                 185                 190

Asp Ile Thr Leu Asn Val Thr Asp His Phe Ala Glu Asn Ala Ala Ile
        195                 200                 205
```

```
Tyr Phe Pro Gln Phe Gly Thr Ala Thr Pro Arg Val Asp Leu Asn Leu
    210                 215                 220

His Arg Met Asn Ala Ser Gln Met Ser Gly Arg Ala Asn Leu Asp Met
225                 230                 235                 240

Cys Leu Tyr Asp Gly Gly Val Lys Ala Arg Ser Leu Gln Met Lys Ile
                245                 250                 255

Glu Gly Ser Asn Lys Ser Gly Thr Gly Phe Gln Val Ile Lys Ser Asp
            260                 265                 270

Ser Ala Asp Thr Ile Asp Tyr Ala Val Ser Met Asn Tyr Gly Gly Arg
        275                 280                 285

Ser Ile Pro Val Thr Arg Gly Val Glu Phe Ser Leu Asp Asn Val Asp
    290                 295                 300

Lys Ala Ala Thr Arg Pro Val Val Leu Pro Gly Gln Arg Gln Ala Val
305                 310                 315                 320

Arg Cys Val Pro Val Pro Leu Thr Leu Thr Thr Gln Pro Phe Asn Ile
                325                 330                 335

Arg Glu Lys Arg Ser Gly Glu Tyr Gln Gly Thr Leu Thr Val Thr Met
            340                 345                 350

Leu Met Gly Thr Gln Thr Pro
        355
```

<210> 6
<211> 151
<212> PRT
<213> Salmonella typhi

<400> 6

```
Met Thr Val Tyr Leu Asp Asp Lys Asp Lys Glu Leu Leu Lys Glu Ile
  1               5                  10                  15

Gln Lys Asp Cys Ala Gln Thr Leu Trp Gln Leu Ala Tyr Lys Val Gly
            20                  25                  30

Leu Thr Pro Thr Pro Cys Phe Lys Arg Leu Lys Lys Leu Lys Asp Arg
        35                  40                  45

Gly Val Ile Ile Gly Gln Phe Ala Leu Leu Asp Lys Glu Lys Leu Gly
    50                  55                  60

Leu Ser Leu Asn Val Phe Ile Met Ile Asn Ile Ser Glu Glu Gln Tyr
65                  70                  75                  80

Ala Ser Ile Ser Glu Lys Ile Lys Ser Met Pro Glu Val Ile Ala Phe
                85                  90                  95

Tyr Arg Ile Ser Gly Ser Phe Asn Tyr Leu Met His Thr Val Phe Thr
            100                 105                 110
```

```
Asp Met Asn Asp Tyr Tyr Ser Phe Tyr Glu Lys Ile Ile Leu Thr Asn
        115                 120                 125

Ser Ser Ile Ser Gly Ser Ala Ser Ser Phe Val Leu Glu Gln Ile Lys
    130                 135                 140

Glu Thr Asn Glu Leu Ser Val
145                 150
```

References:

American Institute of Medicine. (1986) New vaccine development: establishing priorities. Washington, DC: National Academy Press.

Ashcroft, M. T., Ritchie, J. M., Nicholson, C. C. (1964) *Am. J. Hyg.* 79:196-206.

Levine, M. M., Taylor, D. N., Ferreccio, C. (1989) *Pediat. Infect. Dis. J.*, 8:374.

Popoff, M. Y. & Le Minor, L. (1992) *Antigenic formulas of the Salmonella serovars* (WHO Collaborating Center for Reference and Research on *Salmonella*, Institute Pasteur, Paris).

Reeves, M. W., Evins, G. M., Heiba, A. A., Plikaytis, B. D. & Farmer III, J. J. (1989) *J. Clin. Microbiol.* **27**, 313-320.

Rowe, B., Ward, L.R., and Threlfall, E.J. (1997) *Clinical Infectious Disease* 24:(Suppl 1) S106-9

Salyers, A. A. & Whitt, D. D. (1994) *Bacterial Pathogenesis: A molecular approach.* (ASM Press, Washington D.C.).

Yugoslav Typhoid Comission. (1964) *Bull. WHO* 30:623-30.

Claims:

1. Protein encoded by a nucleotide sequence selected from the group consisting of SEQ ID NO 1 and SEQ ID NO 2, or parts thereof, for use in medicine.

2. Antibodies directed against the protein encoded by a nucleotide sequence selected from the group consisting of SEQ ID NO 1 and SEQ ID NO 2, or antigenic fragments thereof for use in medicine.

3. Nucleotide sequence selected from the group consisting of SEQ ID NO 1 and SEQ ID NO 2, or parts thereof, for use in medicine.

4. A vaccine for the protection against diseases caused by *Salmonella enterica* subspecies I, comprising the protein, or parts thereof, encoded by the nucleotide sequence according to SEQ ID NO 1 or antibodies directed against the protein encoded by SEQ ID NO 1, or antigenic fragments thereof and, optionally, a pharmaceutically acceptable carrier.

5. A vaccine for the protection against diseases caused by *Salmonella enterica* subspecies I serovar Typhi, comprising the protein, or parts thereof, encoded by the nucleotide sequence according to SEQ ID NO 2 or antibodies directed against the protein encoded by SEQ ID NO 2, or antigenic fragments thereof and, optionally, a pharmaceutically acceptable carrier.

6. A nucleic acid vaccine for the protection against diseases caused by *Salmonella enterica* subspecies I, comprising SEQ ID NO 1, or parts thereof and, optionally, a pharmaceutically acceptable carrier.

7. A nucleic acid vaccine for the protection against diseases caused by *Salmonella enterica* subspecies I serovar Typhi, comprising SEQ ID NO 2, or parts thereof and, optionally, a pharmaceutically acceptable carrier.

8. A vector vaccine for the protection against diseases caused by *Salmonella enterica* subspecies I, comprising a host in which a recombinant vector comprising SEQ ID NO 1, or parts thereof, has been inserted and, optionally, a pharmaceutically acceptable carrier.

9. A vector vaccine for the protection against diseases caused by *Salmonella enterica* subspecies I serovar Typhi, comprising a host in which a recombinant vector comprising SEQ ID NO 2, or parts thereof, has been inserted and, optionally, a pharmaceutically acceptable carrier.

10. A method for protection against diseases caused by *Salmonella enterica* subspecies I, comprising administering a vaccine according to any of claims 4, 6, and 8.

11. A method for protection against diseases caused by *Salmonella enterica* subspecies I serovar Typhi, comprising administering a vaccine according to any of claims 5, 7, and 9.

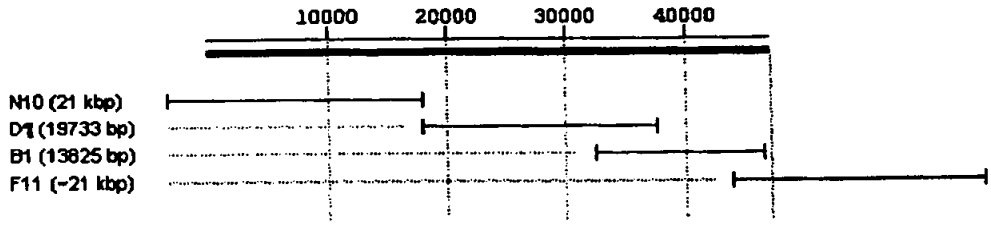
12. Antibodies directed against the protein encoded by a nucleotide sequence selected from the group consisting of SEQ ID NO 1 and SEQ ID NO 2, or antigenic fragments thereof, for use in a diagnostic method.

13. Protein encoded by a nucleotide sequence selected from the group consisting of SEQ ID NO 1 and SEQ ID NO 2, or parts thereof, for use in a diagnostic method.

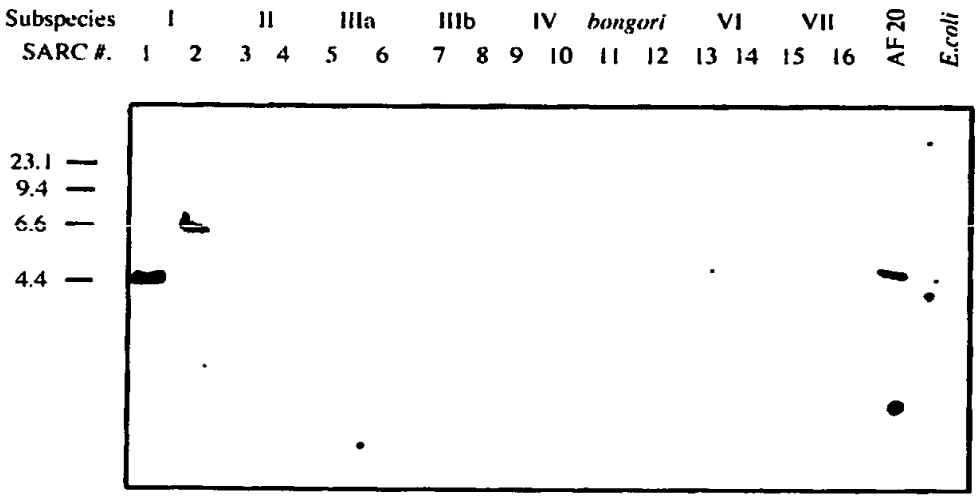
14. Primers for or, probes that hybridize with a nucleotide sequence selected from the group consisting of SEQ ID NO 1 and SEQ ID NO 2, for use in a diagnostic method.

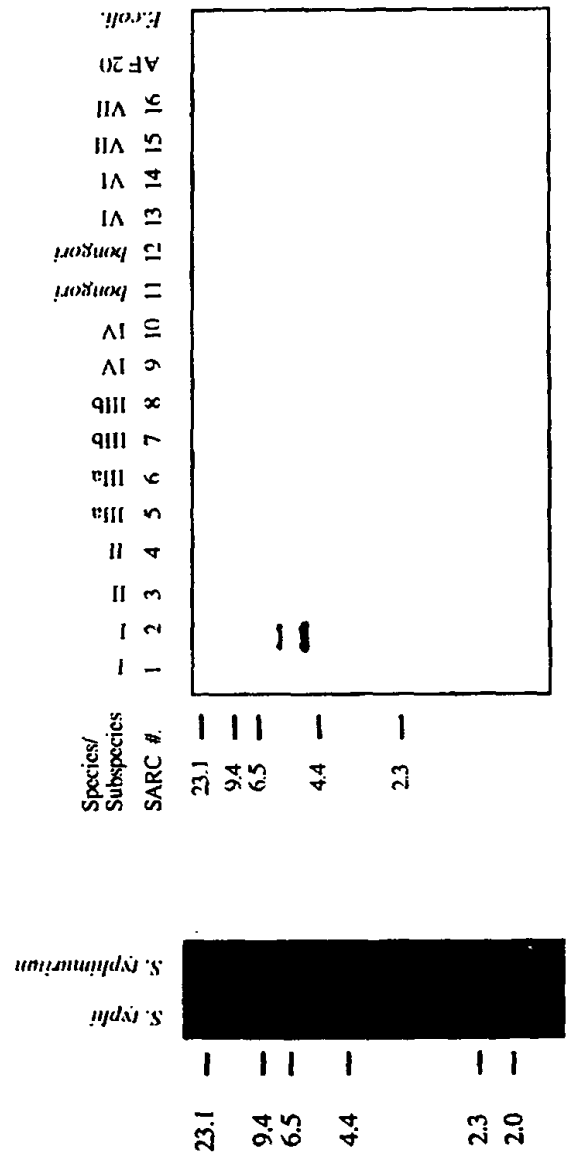
ABSTRACT

The present invention is based on the finding that two fimbrial structures are specific for *Salmonella enterica* subspecies 1 bacteria. Due to their specificity they can be used to provide vaccines against *Salmonella enterica* subspecies I as well as for detection of *Salmonella enterica* subspecies I.


10000
20000
30000
40000
N10 (21 kbp)
D1 (19733 bp)
B1 (13825 bp)
F11 (~21 kbp)

A


Subspecies I II IIIa IIIb IV bongori VI VII AF 20 E.coli
SARC #. 1 2 3 4 5 6 7 8 9 10 11 12 13 14 15 16
23.1
9.4
6.6
4.4


E.coli
AF 20
VII
VII
VI
VI
bongori
bongori
IV
IV
IIIb
IIIb
IIIa
IIIa
II
II
I
I
1 2 3 4 5 6 7 8 9 10 11 12 13 14 15 16
Species/
Subspecies
SARC #.
23.1
9.4
6.5
4.4
2.3

S. typhi
S. typhimurium
23.1
9.4
6.5
4.4
2.3
2.0